



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 109965

TO: Ruixiang Li  
Location: CM1/10D19/10E18  
Art Unit: 1646  
Tuesday, December 16, 2003

Case Serial Number: 09/900448

From: Barb O'Bryen  
Location: Biotech-Chem Library  
CM1-6A05  
Phone: 308-4291 *BOB*

barbara.obryen@uspto.gov

### Search Notes





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 23:41:46 ; Search time 25 Seconds

(without alignments)  
735,498 Million cell updates/sec

Title: US-09-900-448-2

Perfect score: 2185  
Sequence: 1 MARVLGAPVALGWSLCSL.....NAKALPQONTSLGCTH 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt 41:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2139.5	97.9	462	1 HEMO_HUMAN	P02790 homo sapien
2	1700.5	77.8	460	1 HEMO_RABIT	P20058 erythrocytologus
3	1556.5	71.2	460	1 HEMO_RAT	P20059 rattus norv
4	1545.5	70.7	460	1 HEMO_MOUSE	Q91872 mus musculu
5	1544	70.7	459	1 HEMO_PIG	P50828 sus scrofa
6	203	9.3	578	1 MM17_MOUSE	Q91083 mus musculu
7	202.5	9.3	470	1 MM12_HUMAN	P39900 mus sapien
8	199	9.1	606	1 MM17_HUMAN	Q91227 erythrocytologus
9	196	9.0	464	1 MM12_RABIT	P79227 erythrocytologus
10	195.5	8.9	476	1 MM10_MOUSE	O55123 mus musculu
11	195.5	8.9	477	1 MM03_HORSE	Q28397 equus cabal
12	189.5	8.7	476	1 MM01_HUMAN	P09238 homo sapien
13	187	8.6	476	1 MM10_RAT	P07152 rattus norv
14	184.5	8.4	468	1 MM01_RABIT	P13943 erythrocytologus
15	184.5	8.4	492	1 MM11_MOUSE	Q02853 mus musculu
16	183.5	8.4	467	1 MM08_HUMAN	P22894 homo sapien
17	181.5	8.3	469	1 MM01_HUMAN	P03556 homo sapien
18	179.5	8.2	488	1 MM11_HUMAN	P24347 homo sapien
19	177.5	8.1	471	1 MM13_RABIT	O62806 erythrocytologus
20	177	8.1	475	1 MM03_RAT	P03957 rattus norv
21	177	8.1	477	1 MM03_HUMAN	P08254 homo sapien
22	176	8.1	471	1 MM13_HUMAN	P45452 homo sapien
23	176	8.1	472	1 MM13_HORSE	O18327 equus cabal
24	174.5	8.0	469	1 MM01_HORSE	Q93825 equus cabal
25	173	7.9	477	1 MM03_MOUSE	P28862 mus musculu
26	172	7.9	469	1 MM15_HUMAN	P51511 homo sapien
27	169.5	7.8	477	1 MM11_XENLA	Q11004 xenopus lae
28	169.5	7.8	478	1 VTNC_HUMAN	P04004 homo sapien
29	167.5	7.7	508	1 MM19_HUMAN	Q95942 mus sapien
30	166.5	7.6	466	1 MM13_RAT	P23097 rattus norv
31	166	7.6	469	1 MM01_PIG	P21592 sus scrofa
32	166	7.6	478	1 MM03_RABIT	P28863 erythrocytologus
33	165.5	7.6	478	1 VTNC_MOUSE	P29788 mus musculu

34	165	7.6	657	1 MM15_MOUSE	O54732 mus musculu
35	163	7.5	618	1 MM24_MOUSE	Q91082 mus musculu
36	163	7.5	618	1 MM24_RAT	Q99966 rattus norv
37	162.5	7.4	645	1 MM24_HUMAN	Q91872 mus sapien
38	161.5	7.4	465	1 MM08_MOUSE	O87013 mus musculu
39	159.5	7.3	466	1 MM08_RAT	O88766 rattus norv
40	158	7.2	459	1 VTNC_PIG	P48819 sus scrofa
41	155.5	7.1	467	1 MM18_XENLA	O13065 xenopus lae
42	154.5	7.1	472	1 MM13_MOUSE	P33435 mus musculu
43	152	7.0	607	1 MM16_HUMAN	P51512 homo sapien
44	151	6.9	469	1 MM01_BOVIN	P28053 bos taurus
45	151	6.9	607	1 MM16_MOUSE	Q91870 mus musculu

## ALIGNMENTS

RESULT 1	ID	HEMO_HUMAN	STANDARD;	PRT;	462 AA.
AC	P02790;	21-VUL-1986 (Rel. 01, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Hemopexin precursor (Beta-1B-glycoprotein).				
CN	HPX.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88316972; PubMed=2842511; Silengo L.;				
RA	Altuda F., Poli V., Rescagno G., Silengo L.;				
RT	"Structure of the human hemopexin gene and evidence for				
RT	intron-mediated evolution."				
RL	J. Mol. Evol. 27:102-108 (1988).				
RN	[2]				
RP	SEQUENCE OF 2-462 FROM N.A.				
RX	MEDLINE=89122012; PubMed=3220477;				
RA	Law M.L., Cai G.Y., Hartz J.A., Jones C., Kao P.T.;				
RT	"The hemopexin gene maps to the same location as the beta-globin gene				
RL	cluster on human chromosome 11."				
RN	Genomics 3:48-52 (1988).				
RP	SEQUENCE OF 22-462 FROM N.A.				
RX	MEDLINE=85242073; PubMed=2989777;				
RA	Altuda F., Poli V., Rescagno G., Argos P., Cortese R., Silengo L.;				
RT	"The primary structure of human hemopexin deduced from cDNA sequence:				
RT	evidence for internal, repeating homology."				
RL	Nucleic Acids Res. 13:3841-3859 (1985).				
RN	[4]				
RP	SEQUENCE OF ACTIVE PROTEIN.				
RX	MEDLINE=85113173; PubMed=3855550;				
RA	Takahashi N., Takahashi Y., Putnam F.W.;				
RT	"Complete amino acid sequence of human hemopexin, the heme-binding				
RT	protein of serum."				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:73-77 (1985).				
RN	[5]				
RP	SEQUENCE OF 24-255.				
RX	MEDLINE=85076955; PubMed=6510521;				
RA	Frantkova V., Borvak J., Klus I., Moravsek L.;				
RT	"Amino acid sequence of the N-terminal region of human hemopexin."				
RL	FEBS Lett. 178:213-216 (1984).				
RN	[6]				
RP	PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.				
RX	MEDLINE=84193947; PubMed=6371807;				
RA	Takahashi N., Takahashi Y., Putnam F.W.;				
RT	"Structure of human hemopexin: O-glycosyl and N-glycosyl sites and				
RT	unusual clustering of tryptophan residues."				
RL	Proc. Natl. Acad. Sci. U.S.A. 81:2021-2025 (1984).				
CC	-I- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN				
CC	AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE				

CIRCULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M36803; AA58678.1; -  
 DR EMBL: M36796; AA58678.1; JOINED.  
 DR EMBL: M36799; AA58678.1; JOINED.  
 DR EMBL: M36800; AA58678.1; JOINED.  
 DR EMBL: M36801; AA58678.1; JOINED.  
 DR EMBL: M36802; AA58678.1; JOINED.  
 DR EMBL: M36803; AA58678.1; JOINED.  
 DR EMBL: X02537; CA26382.1; ALT\_INIT.  
 DR PIR: I56456; OOHU.  
 DR HSSP: P20058; 1HXN.  
 DR SWISS-2DPAGE: P02790; HUMAN.  
 DR Siema-2DPAGE: P02790; -  
 DR Genew: HGNC:5171; HPX.  
 DR MIM: 142290; -  
 DR GO: GO:0005615; C:extracellular space; TAS.  
 DR GO: GO:0005488; F:binding activity; TAS.  
 DR GO: GO:0015332; F:heme transporter activity; TAS.  
 DR GO: GO:0005211; F:plasma glycoprotein; TAS.  
 DR GO: GO:0015886; F:heme transporter; TAS.  
 DR GO: GO:0006879; F:iron ion homeostasis; TAS.  
 DR InterPro: IPR000585; Hemopexin.  
 DR Pfam: PF00045; hemopexin; 5.  
 DR SMART: SM00120; HX; 5.  
 DR PROSITE: PS00024; HEMOPEXIN; 2.  
 DR GlycoProtein: Heme; Plasma; Repeat; Transport; Signal.  
 KW SIGNA 1 23  
 FT CHAIN 24 462  
 FT DOMAIN 56 93 HEMOPEXIN-LIKE 1.  
 FT DOMAIN 97 141 HEMOPEXIN-LIKE 2.  
 FT DOMAIN 188 231 HEMOPEXIN-LIKE 3.  
 FT DOMAIN 263 306 HEMOPEXIN-LIKE 4.  
 FT DOMAIN 308 351 HEMOPEXIN-LIKE 5.  
 FT METAL 79 79 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT METAL 150 150 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT DISULFID 50 231  
 FT DISULFID 149 154  
 FT DISULFID 188 200  
 FT DISULFID 257 460  
 FT DISULFID 366 408  
 FT DISULFID 418 435  
 FT CARBOHYD 24 24  
 FT CARBOHYD 64 64  
 FT CARBOHYD 187 197  
 FT CARBOHYD 240 240  
 FT CARBOHYD 246 246  
 FT CARBOHYD 453 453  
 FT SEQUENCE 462 AA; 51676 MW; 054844D0603763B8 CRC64;  
 Query Match 97.9%; Score 2139.5; DB 1; Length 462;  
 Best Local Similarity 84.6%; Pred. No. 5.8e-176;  
 Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;  
 QY 1 MARVLGAPVALGLMSLCMSLAITPLPTSAHGNVAGETKPPDPDYTERGSDGMSFPDATT 60  
 DB 1 MAVLGAPVALGLMSLCMSLAITPLPTSAHGNVAGETKPPDPDYTERGSDGMSFPDATT 60  
 QY 61 LDDNGTMLFFKSGFVWKSFKWDELLISERWKNFPPSDAAFRQGHNSVFLIKGDKVWVVP 120  
 DB 61 LDDNGTMLFFKSGFVWKSFKWDELLISERWKNFPPSDAAFRQGHNSVFLIKGDKVWVVP 120

QY 121 PEKKEGYPKLQDEPFPGIPSPPLDAAVECHRECOAGVLFQ----- 163  
 DB 121 PEKKEGYPKLQDEPFPGIPSPPLDAAVECHRECOAGVLFQGDENWMTLATGTMKER 180  
 QY 164 -----GHGRN 169  
 DB 181 SMPAVNGSSALRWLGRYYCFQGNQPLRFPDVRGVEPRYPYRDVDFMPCPGRGHNRN 240  
 QY 170 GTCHGNSHTHGPRYMCSPHLVLSALTSNDHGTVPFSGTHWRRLTSDGMSWPIAHQ 229  
 DB 241 GTCHGNSHTHGPRYMCSPHLVLSALTSNDHGTVPFSGTHWRRLTSDGMSWPIAHQ 300  
 QY 230 WPGSPAVDAAAFSWEKLYLVQSTQYVYFLTKGTYLVGYPRLKEKVGTPHGIILDSV 289  
 DB 301 WPGSPAVDAAAFSWEKLYLVQSTQYVYFLTKGTYLVGYPRLKEKVGTPHGIILDSV 360  
 QY 290 DAAFCPGSSRLHIMAGRRLWDLKSGAQTWELPWEKVDGALCKEKSIGPNSCSA 349  
 DB 361 DAAFCPGSSRLHIMAGRRLWDLKSGAQTWELPWEKVDGALCKEKSIGPNSCSA 420  
 QY 350 NGPGLYLHGPNIYCYSDVEKLNAAKALPQPQNVTSLLGCTH 391  
 DB 421 NGPGLYLHGPNIYCYSDVEKLNAAKALPQPQNVTSLLGCTH 462  
 RESULT 2  
 ID HEMO\_RABBIT STANDARD; PRT; 460 AA.  
 AC P20058;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hemopexin precursor.  
 GN HPX.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxId=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND HEME-BINDING SITES.  
 RC TISSUE=Liver;  
 RX MEDLINE=93203213; PubMed=7681064;  
 RA Morgan W.T., Master P., Tatum P., Kao S.-M., Alam J., Smith A.;  
 RT "Identification of the histidine residues of hemopexin that  
 RT coordinate with heme-iron and of a receptor-binding region";  
 RL J. Biol. Chem. 268:6256-6262(1993).  
 RN [2]  
 RP SEQUENCE OF 26-53.  
 RX MEDLINE=88339942; PubMed=3421961;  
 RA Wellner D., Cheng K.C., Mueller-Eberhard U.;  
 RT "N-terminal amino acid sequences of the hemopexins from chicken, rat  
 RT and rabbit";  
 RL Biochem. Biophys. Res. Commun. 155:622-625(1988).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 242-460.  
 RX MEDLINE=96173004; PubMed=8590016;  
 RA Faber H.R., Groom C.R., Baker H.M., Morgan W.T., Smith A., Baker E.N.;  
 RT "1.8-A crystal structure of the C-terminal domain of rabbit serum  
 RT haemopexin";  
 RL Structure 3:551-559(1995).  
 CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN  
 CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE  
 CC CIRCULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M62642; AAA4137.1; -  
 CC EMBL: X60006; CAA42621.1; -  
 CC PIR: A43079; OORT.  
 CC HSSP: P20058; 1HXN.  
 CC InterPro: IPR000585; Hemopexin.  
 CC Pfam: PF00045; hemopexin; 5.  
 CC SMART: SM00120; HX; 5.  
 CC PROSITE: PS00024; HEMOPEXIN; 1.  
 CC GlycoProtein; Heme; Plasma; Repeat; Transport; Signal.  
 CC  
 CC FT CHAIN 1 23  
 CC FT SIGNAL 1 23  
 CC FT DOMAIN 24 460 HEMOPEXIN-LIKE 1.  
 CC FT DOMAIN 56 93 HEMOPEXIN-LIKE 2.  
 CC FT DOMAIN 97 140 HEMOPEXIN-LIKE 3.  
 CC FT DOMAIN 187 230 HEMOPEXIN-LIKE 4.  
 CC FT DOMAIN 261 304 HEMOPEXIN-LIKE 5.  
 CC FT DOMAIN 306 349 HEMOPEXIN-LIKE 5.  
 CC FT METAL 79 79 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 CC FT METAL 149 149 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 CC FT DISULFID 50 230 BY SIMILARITY.  
 CC FT DISULFID 148 153 BY SIMILARITY.  
 CC FT DISULFID 187 199 BY SIMILARITY.  
 CC FT DISULFID 255 458 BY SIMILARITY.  
 CC FT DISULFID 364 406 BY SIMILARITY.  
 CC FT DISULFID 416 433 BY SIMILARITY.  
 CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CONFLICT 38 38 N-> C (IN REF. 2).  
 CC FT CONFLICT 49 50 HC-> KW (IN REF. 2).  
 CC FT SEQUENCE 460 AA; 51291 MW; A1D67D05B5BFD83 CRC64;  
 CC  
 CC Query Match 71.2%; Score 1556.5; DB 1; Length 460;  
 CC Best Local Similarity 63.8%; Pred. No. 5.8e-126; Indels 79; Gaps 6;  
 CC Matches 296; Conservative 35; Mismatches 54;  
 CC  
 CC QY 1 MARVLGAPVALGLMSLCWSLAATPLPTSAHGNVAEGE--TRPDPYTERCSGDSGSPDA 58  
 CC DB 1 MARTVALNITLVILGLCWSLAVANPLP--AAHETVAKGNGTGRPDSVDVIEHCSDAMSFDA 58  
 CC QY 59 TITLDNGTMLFFPGGEFVWVSKMDRELISERWNPSPVDAAFRQGHNSVFLIKDKVWY 118  
 CC DB 59 TITLDNGTMLFFPGGEFVWVSKMDRELISERWNPSPVDAAFRQGHNSVFLIKDKVWY 117  
 CC QY 119 YPEKKEKGVKPLQDEBPGIPSPDLAAVCHGEGCOABGVFFPOCH-- 165  
 CC DB 118 YPEKKEKGVKPLQDEBPGIPSPDLAAVCHGEGCOABGVFFPOCH-- 165  
 CC QY 166 -----GH-- 167  
 CC DB 166 -----GH-- 167  
 CC QY 178 EBSWPAVNGCTAALRWLERVYFCQGNKFLRFPNVTGEVPRRYLDARDYFISCPGKHGK 217  
 CC DB 178 EBSWPAVNGCTAALRWLERVYFCQGNKFLRFPNVTGEVPRRYLDARDYFISCPGKHGK 217  
 CC QY 168 RNGTGHGNSYTHHGPEYMRCSPLVLALTSNKGATVAESGTHYWRDLTSDRGHWSPI 226  
 CC DB 238 LNRGTGHGNSYTHHGPEYMRCSPLVLALTSNKGATVAESGTHYWRDLTSDRGHWSPI 225  
 CC QY 227 AHQWPGGSAVDAAFESWBEKLYIVOSTGVYVFLTKGTYLVSGYPRKLEKVEVTPHGILL 286  
 CC DB 296 AHQWPGGSAVDAAFESWBEKLYIVOSTGVYVFLTKGTYLVSGYPRKLEKVEVTPHGILL 286  
 CC QY 287 DSVDAAFICPGSSRLIMAGRLMWLDLXSGAQTATELFWPEKVDGALCKEKSIGSPS 346  
 CC DB 287 DSVDAAFICPGSSRLIMAGRLMWLDLXSGAQTATELFWPEKVDGALCKEKSIGSPS 346

DB 356 DTDIAAFSCPGSSKLYVTSGRLMWLDLXSGAQTAAELSWPEKVDGALCKEKSIGSPS 415  
 QY 347 CSANGGLYLIRGPNLYCYSDVEKMAAKALPPQVNTSLIGCT 390  
 DB 416 CSANGGLYLIRGPNLYCYSDVEKMAAKALPPQVNTSLIGCT 459  
 RESULT 4  
 HEMO MOUSE  
 ID HEMO MOUSE STANDARD; PRT; 460 AA.  
 AC Q91X72; P97824; Q8WUP0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hemopexin precursor.  
 GN HPX OR HPXN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan E., Moore T., Max S.I., Wang J., Heich F.,  
 RA Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Pangue C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mollath S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 7-460 FROM N.A.  
 RC TISSUE=Liver;  
 RA Koepfel R.R., Rohrbach D.H., Breckeliser B.B.;  
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN  
 CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE  
 CC CIRCULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: BC011246; AAH11246.1; -  
 CC EMBL: BC019901; AAH19901.1; -  
 CC EMBL: U89889; AAB49490.1; -  
 CC MGD: MG1:105112; Hpxn.  
 CC InterPro: IPR000585; Hemopexin.  
 CC Pfam: PF00045; hemopexin; 5.  
 CC SMART: SM00120; HX; 5.  
 CC PROSITE: PS00024; HEMOPEXIN; 1.



**THIS PAGE BLANK (user)**

LOCUS	AV653336	631 bp	mRNA	linear	EST 15-JAN-2002
DEFINITION	AV653336	GLC Homo sapiens cDNA clone GLCDUH02 3', mRNA sequence.			
ACCESSION	AV653336				
VERSION	AV653336.1	GI:9874350			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 631) Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,O., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Zhong,W., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001) 21625106 11752456				
JOURNAL	CONTACT: Zeguang Han				
MEDLINE	Chinese National Human Genome Center at Shanghai				
PUBMED	351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.				
FEATURES	Location/Qualifiers				
SOURCE	1..631 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="GLCDUH02" /tissue_type="corresponding non cancerous liver tissue" /dev_stage="Adult" /lab_host="SOLR" /clone_lib="GLC" /note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT	138 a 180 c 161 g 149 t 3 others				
ORIGIN	XhoI"				
Query Match	11.4%; Score 425.6; DB 9; Length 631;				
Best Local Similarity	98.4%; Pred. No. 1.1e-99;				
Matches 428; Conservative	7; Mismatches 7; Indels 0; Gaps 0;				
1479	CAGGACGGCGGCTGTGTGTCGTCGACCGTGAAGTCAAGAGCCCAAGCCAGTGGACAGAGC	1538			
Db	106 CAGACGGCGGCTGTGTGTCGTCGACCGTGAAGTCAAGAGCCCAAGCCAGTGGACAGAGC	165			
1539	TTCCCTGGCCCCCATGAGAAAGGTAGACGAGCCCTGTGTATGAGAAAGTCCCTGGCCCTTA	1598			
Db	166 TTCCTGGCCCCCATGAGAAAGGTAGACGAGCCCTGTGTATGAGAAAGTCCCTGGCCCTTA	225			
1599	ACTCATGTTCCGGCCCAATGATGTCCCGGCTGTACTCATCTCATAGTGTCCCAATTTGACTGCT	1658			
Db	226 ACTCATGTTCCGGCCCAATGATGTCCCGGCTGTACTCATCTCATAGTGTCCCAATTTGACTGCT	285			
1659	ACAGTGAATGTGAGAAACTGAATGACGACCAAGGCCCTTCGCGCAACCCAGAGATGACCA	1718			
Db	286 ACAGTGAATGTGAGAAACTGAATGACGACCAAGGCCCTTCGCGCAACCCAGAGATGACCA	345			
1719	GTCCTCTGGGCTGCACTCATGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTC	1778			
Db	346 GTCCTCTGGGCTGCACTCATGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTC	405			
1779	CTAGTCTCTAATAATTAAGACAGATGTGTTCTGTGCTCTCACTGAGGGGCTTCTGCA	1838			
Db	406 CTAGTCTCTAATAATTAAGACAGATGTGTTCTGTGCTCTCACTGAGGGGCTTCTGCA	465			

	Dy	1839	TGAGTCGGCGCTGGCCCCACCTCCCAAGTTTCTCATATATAAAGCAGAATTGCCTTCCAC	1898
	Dd	466	TGAactTGcctgcccacccaccacctccccagtttctcatatataaaagaattgcctttcac	525
OY		1899	TTGAAATCAGGAGACC 1913       	
	RESULT 7			
	AA705094/c		548 bp    mRNA       linear   EST 24-DEC-1997	
	LOCUS	c2193n10.gI Soares fetal liver_infls.SI Homo sapiens CDNA		
	DEFINITION	clone IMAGE:462493_3' similar to gb:f03048 HEMOPEXIN PRECURSOR (HUMAN) ; mRNA sequence.		
	ACCESSION	AA705094		
	VERSION	AA705094.1 GI:2715012		
	KEYWORDS	EST.		
	SOURCE	Homo sapiens (human)		
	ORGANISM	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homniidae; Homo.		
	REFERENCE	1 (bases 1 to 548) Hiller L., Allen M., Bowles J., Dubugue T., Geisel G., Joet S., Kitzman D., Kucba T., Lucy M., Le N., Lennon G., Marra M., Martin U., Moore B., Schellenberg R., Stepeco M., Tan P., Theising B., White Y., Wylie T., Waterston R. and Wilson R. WashU-NCI human EST project Unpublished		
	AUTHORS	Contact: Wilton RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watscn.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information. Seq primer:-40mJ3 fwd, ET from Amersharn High quality sequence stop: 462. Location/Qualifiers		
FEATURES				
source		1..548	/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="GBD:3753522"	
			/db_xref="taxon:9606"	
			/clone="IMAGE:462493"	
			/sex="male"	
			/dev_stage="20 week-post conception fetus"	
			/lab_host="DH10B (ampicillin resistant)"	
			/clone_id="Soares fetal liver spleen_INFLS_S1"	
			/note="Organ: Liver and Spleen; Vector: pTV73D (pharmacia) with a modified polylinker; Site 1: Pac I; site 2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INPLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5]. AACTGCAAGATTAATTAAGAATCTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."	
BASE COUNT		137 a      137 c      160 g      113 t          1 others		
ORIGIN				
		Query Match             .11.3%; Score 423; DB 9; Length 548; Best Local Similarity 100.0%; Pred. NO.4.7e-99; Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY		1479	CAGCACGCCGCTGTGTGGCTGGACCTTGAAGTCAGAGCCCMA GCCAGTGA CAGAGC	1538
Dd		423	CAGACGCGCGCTGTGTGGCTGGACCTTGAAGTCAGAGCCCMA GCCAGTGA CAGAGC	364
OY		1539	TYCTTGggccccaTGAgAggtAcGgaGcCTTtgTgTaTgAAAAAGtcCTTgGccCTta	1598

```

Db      363  TTCTTGGCCCAAGAGTAGACGAGCTTGTATAGAAAGTCCCTTGGCCCTA 304
Qy      1599 ACTCATGTCGGCCATGATGTCCTGTAACATCATCAATGATTCCTTACTGCT 1658
Db      303  ACTCATGTCGGCCATGATGTCCTGTAACATCATCAATGATTCCTTACTGCT 244
Qy      1659 ACAAGTATGAGAGAACTGATGACAGCCAGGCTTCGCAACCCAGATGTGACCA 1718
Db      243  ACAAGTATGAGAGAACTGATGACAGCCAGGCTTCGCAACCCAGATGTGACCA 184
Qy      1719 GTCTCCCTGGGCTGACACTGAGGGGCTTTCATGATGATGCTGCTGGCCCACTC 1778
Db      183  GTCTCCCTGGGCTGACACTGAGGGGCTTTCATGATGATGCTGCTGGCCCACTC 124
Qy      1779 CTAGTTCCTCATATATAAGACAGATGCTTCTGCTTCTCACTGAGGGGCTTTCACA 1838
Db      123  CTAGTTCCTCATATATAAGACAGATGCTTCTGCTTCTCACTGAGGGGCTTTCACA 64
Qy      1839 TGAAGTCTGGCTGGCCCACTCCAGTTTCTCATATAAAGACAGATGCTTCTCAC 1898
Db      63  TGAAGTCTGGCTGGCCCACTCCAGTTTCTCATATAAAGACAGATGCTTCTCAC 4
Qy      1899 TTG 1901
Db      3  TTG 1

RESULT 8
A1798878 562 bp mRNA linear EST 18-DEC-1999
LOCUS     w693908.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2348702 3' similar to gb:U03048 HEMOPLEXIN PRECURSOR (HUMAN) ;,
mRNA sequence.
A1798878
VERSION   A1798878.1 GI:5364350
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 562)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1426 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 465.
Location/Qualifiers
1. 562
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2348702"
/lab_host="DH10B"
/clone_id="Soares_NFL_T_GBC_S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung Mh19w, testis NHT, and B-cell
NCI-CCAP GCBI) were mixed, and 88 circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

```

BASE COUNT 134 a 148 c 161 g 119 t
ORIGIN
Query Match 11.2%; Score 419.8; DB 9; Length 562;
Best Local Similarity 99.5%; Pred. No. 3,3e-98;
Matches 421; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1479 CAGAGCGCGCTGTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGATGTGACAGC 1538
Db      423  CAGAGCGCGCTGTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGATGTGACAGC 364
Qy      1539 TTCTTGGCCCAAGTGAAGATGAGAGCTTGTGTATGAAAGTCCCTTGGCCCTA 1598
Db      363  TTCTTGGCCCAAGTGAAGATGAGAGCTTGTGTATGAAAGTCCCTTGGCCCTA 304
Qy      1599 ACTCATGTCGGCCATGATGTCCTGTAACATCATCAATGATTCCTTACTGCT 1658
Db      303  ACTCATGTCGGCCATGATGTCCTGTAACATCATCAATGATTCCTTACTGCT 244
Qy      1659 ACAAGTATGAGAGAACTGATGACAGCCAGGCTTCGCAACCCAGATGTGACCA 1718
Db      243  ACAAGTATGAGAGAACTGATGACAGCCAGGCTTCGCAACCCAGATGTGACCA 184
Qy      1719 GTCTCCCTGGGCTGACACTGAGGGGCTTTCATGATGATGCTGCTGGCCCACTC 1778
Db      183  GTCTCCCTGGGCTGACACTGAGGGGCTTTCATGATGATGCTGCTGGCCCACTC 124
Qy      1779 CTAGTTCCTCATATATAAGACAGATGCTTCTGCTTCTCACTGAGGGGCTTTCACA 1838
Db      123  CTAGTTCCTCATATATAAGACAGATGCTTCTGCTTCTCACTGAGGGGCTTTCACA 64
Qy      1839 TGAAGTCTGGCTGGCCCACTCCAGTTTCTCATATAAAGACAGATGCTTCTCAC 1898
Db      63  TGAAGTCTGGCTGGCCCACTCCAGTTTCTCATATAAAGACAGATGCTTCTCAC 4
Qy      1899 TTG 1901
Db      3  TTG 1

RESULT 9
AV655383 433 bp mRNA linear EST 15-JAN-2002
LOCUS     AV655383 GLC Homo sapiens cDNA clone G15GAL10 3', mRNA sequence.
DEFINITION AV655383
VERSION   AV655383.1 GI:9876397
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 433)
Xiao,H., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,W., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 433
/organism="Homo sapiens"

```







double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pUT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 99 a 107 c 123 g 89 t 1 others

Query Match 11.0%; Score 412; DB 14; Length 419;  
Best Local Similarity 99.8%; Pred. No. 3.1e-96;  
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1479 CAGACGCGCGCTGTGGTGGCTGACCTGAGTCAAGAGCCCAAGCCAGTGAAGAGC 1538  
413 CAGACGCGCGCTGTGGTGGCTGACCTGAGTCAAGAGCCCAAGCCAGTGAAGAGC 354  
1539 TTCCTTGGCCCAATGAGAGTGAAGAGCCCTTGTGTATGAAAAGTCCCTTGGCCCTA 1598  
353 TTCCTTGGCCCAATGAGAGTGAAGAGCCCTTGTGTATGAAAAGTCCCTTGGCCCTA 294  
1599 ACTCATGTTCCGCGCATGATGCTCCGCGCTTGTACCTCATCCATGATGCTCCCAATTTGTACTGCT 1658  
293 ACTCATGTTCCGCGCATGATGCTCCGCGCTTGTACCTCATCCATGATGCTCCCAATTTGTACTGCT 234  
1659 ACAGATGATGAGAACTGATGACCAAGGCGCTTCCGCAACCCAGATGTGACCA 1718  
233 ACAGATGATGAGAACTGATGACCAAGGCGCTTCCGCAACCCAGATGTGACCA 174  
1719 GTCTCTGGGCTGCTGCTGAGAGGCGCTTGTGACATGATGCTGAGCGCCCACTC 1778  
173 GTCTCTGGGCTGCTGCTGAGAGGCGCTTGTGACATGATGCTGAGCGCCCACTC 114  
1779 CTAGTTCCTCATATAAAGACAGATGCTTCTTGTGCTTCTGACAGAGGCGCTTGTGACA 1838  
113 CTAGTTCCTCATATAAAGACAGATGCTTCTTGTGCTTCTGACAGAGGCGCTTGTGACA 54  
1839 TGAGTCTGGGCTGAGCCCACTCCCGAGTTCTCATATAAAGACAGATGCT 1891  
53 TGAGTCTGGGCTGAGCCCACTCCCGAGTTCTCATATAAAGACAGATGCT 1

RESULT 13 1053 bp mRNA linear EST 31-MAY-2003  
LOCUS AL564329 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
DEFINITION CS0DM004YJ20 3-PRIME, mRNA sequence.  
ACCESSION AL564329  
VERSION AL564329.2 GI:31288307  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1053)  
Li, W.B., Gruber, C., Jesse, J. and Polayres, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 16, 2001 this sequence version replaced gi:12914625.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DM004D10NP1&cluster=5958.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DM004D10NP1.  
Location/Qualifiers  
1. 1053  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM004YJ20"  
/issue="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/note="Organ: liver; Vector: PCWSPORT 6; 1st strand cDNA  
was primed with a NotI-soligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the PCWSPORT 6  
vector. Library was not normalized."

BASE COUNT 237 a 286 c 306 g 210 t 14 others

Query Match 11.0%; Score 411.6; DB 9; Length 1053;  
Best Local Similarity 72.3%; Pred. No. 6.3e-96;  
Matches 698; Conservative 2; Mismatches 11; Indels 254; Gaps 3;

928 TCACAGGAGACCACTACTGCGCTGACACACGCGGGATGGCTGATAGTGGCC 987  
719 TTCAAGGAGACCACTACTGCGCTGACACACGCGGGATGGCTGATAGTGGCC 660  
988 ATTGCTCATAGTGGCCCAAGGCTCTTCAAGATGATGCTCTTCTTCTGGAAGAA 1047  
659 ATTGCTCATAGTGGCCCAAGGCTCTTCAAGATGATGCTCTTCTTCTGGAAGAA 600  
1048 AAATCTATCTGCTCCAGGTGTGTATTGGGGAGAGGCTTGAAGTGAAGACTGGACAA 1107  
599 AAATCTATCTGCTC----- 585  
1108 CATATCAACTGTGATTATTACATCTTGTCTCCAGGGACCCAGATATATGCT 1167  
584 -----CAGGACCCAGATATATGCT 563  
1168 TCCTGACAAAGGAGGCTATACCTAGTAAAGGCTTATCCGAAGCGCTGAGAGGAAG 1227  
562 TCCTGACAAAGGAGGCTATACCTAGTAAAGGCTTATCCGAAGCGCTGAGAGGAAG 503  
1228 TCGGAGCCCTCATGAGGATTTATCTGAGCTCTGTGATGCGGCTTTATCTGCGGT 1287  
502 TCGGAGCCCTCATGAGGATTTATCTGAGCTCTGTGATGCGGCTTTATCTGCGGT 443  
1288 CTTCTGCGCTCATATCATGAGGAGGAGGCTTGTGGGTGCTTGAAGGCGACTTTGT 1347  
442 CTTCTGCGCTCATATCATGAG----- 422  
1348 CTGCTACTGTCTGTGGCATAGTATCCCAACAGAGGATGAGAGGCTAGTCAAGATCC 1407  
421 ----- 422  
1408 CCAGGGCATGAGAGGCTTAGTCAAGATCCCATGATGAGAGCATATATGTTGG 1467  
421 ----- 422  
1468 TGCTTTTCCCAAGAGCGCGGTGTGTGAGCTGAGCTGAGTCAAGAGCCCAAGCCAC 1527  
421 -----CAGGACGCGCGGTGTGTGAGCTGAGCTGAGTCAAGAGCCCAAGCCAC 373  
1528 GTGACAGAGCTTCTTGGCCCATGAGAGGATGAGAGGCTTGTATGAGAAAGTC 1587  
372 GTGACAGAGCTTCTTGGCCCATGAGAGGATGAGAGGCTTGTATGAGAAAGTC 313  
1588 CTTGGCCCTTAATCATGTTTCCGCAATGATGCTCCGCTTGTATCATCATGATGCTCCAA 1647  
312 CTTGGCCCTTAATCATGTTTCCGCAATGATGCTCCGCTTGTATCATCATGATGCTCCAA 253  
1648 TTGTATCTGCTACAGTATGAGAGAACTGATGAGCAAGGCGCTTCCGCAAGCCCA 1707  
252 TTGTATCTGCTACAGTATGAGAGAACTGATGAGCAAGGCGCTTCCGCAAGCCCA 193  
1708 GAATGACCAAGTCTCTGAGGCTGACATGAGAGGCGCTTCTGACATGATGCTGGCT 1767  
192 GAATGACCAAGTCTCTGAGGCTGACATGAGAGGCGCTTCTGACATGATGCTGGCT 133

QY 1768 GGGCCCACTCTAGTTCCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGG 1827  
 DB 132 GGGCCCACTCTAGTTCCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGG 73  
 QY 1828 GCTTCTGACATGATGCTGCGCTTGCGCCACCTCC-CGAGTTTCTCATATAAAGACAGA 1886  
 DB 72 GCTTCTGACATGATGCTGCGCTTGCGCCACCTCC-CGAGTTTCTCATATAAAGNNNA 13  
 QY 1887 TTGCT 1891  
 DB 12 WNWMT 8  
 RESULT 14  
 AA514026 485 bp mRNA linear EST 13-NOV-1997  
 LOCUS n132e01.s1 NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:972024  
 DEFINITION similar to gb:J03048 HEMOPEXIN PRECURSOR (HUMAN) ; mRNA sequence.  
 ACCESSION AA514026  
 VERSION AA514026.1 GI:2252447  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 485)  
 NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-@email.nih.gov  
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D. Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.lnl.gov/bdrip/image/image.html  
 Insert Length: 602 Std Error: 0.00  
 Seq primer: -40ml3 fwd. RT from Amersham  
 High quality sequence stop: 388.  
 FEATURES  
 source  
 1.485  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:972024"  
 /tissue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Ew1"  
 /note="Vector: PAMPI0; mRNA made from Ewing's sarcoma,  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size-selected on agarose gel, average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."  
 BASE COUNT 96 a 143 c 124 g 122 t  
 ORIGIN  
 Query Match 11.0%; Score 411; DB 9; Length 485;  
 Best Local Similarity 99.8%; Pred. No. 6,1e-96;  
 Matches 422; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1479 CAGGACGGCGCTGTGTGCTGAGCTGAGCTGAGGCCCAAGCTGAGCAGAGC 1538  
 DB 64 CAGGACGGCGCTGTGTGCTGAGCTGAGCTGAGGCCCAAGCTGAGCAGAGC 123  
 QY 1539 TTCTCTGGCCCATGAGAGGTAGACGAGCCTTGTATGAAAAGTCCCTTGCCCTTA 1598  
 DB 124 TTCTCTGGCCCATGAGAGGTAGACGAGCCTTGTATGAAAAGTCCCTTGCCCTTA 183

QY 1599 ACTCATGTTCCGCAATGATGCTCCGCGCTGTACTCATTCATGATGCTCCCAATTGTACTGCT 1658  
 DB 184 ACTCATGTTCCGCAATGATGCTCCGCGCTGTACTCATTCATGATGCTCCCAATTGTACTGCT 243  
 QY 1659 ACAATGATGTGGAGAACTGAAATGCAAGCCAGGCGCTTCCGCAACCCAGAAATGTGACCA 1718  
 DB 244 ACAATGATGTGGAGAACTGAAATGCAAGCCAGGCGCTTCCGCAACCCAGAAATGTGACCA 303  
 QY 1719 GTCTCCGCGGCTCACTCACTGAGGGGCGCTTGTGACATGATGCTGAGCGGCGCCACCTC 1778  
 DB 304 GTCTCCGCGGCTCACTCACTGAGGGGCGCTTGTGACATGATGCTGAGCGGCGCTTGTGAC 362  
 QY 1779 CTAGTTCCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGGCGCTTGTGAC 1838  
 DB 363 CTAGTTCCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGGCGCTTGTGAC 422  
 QY 1839 TGAATCTGCGCTTGCGCCACCTCCCAAGTTTCTCATATAAAGACAGATTGCTTCTTAC 1898  
 DB 423 TGAATCTGCGCTTGCGCCACCTCCCAAGTTTCTCATATAAAGACAGATTGCTTCTTAC 482  
 QY 1899 TTG 1901  
 DB 483 TTG 485  
 RESULT 15  
 A1283497/c 494 bp mRNA linear EST 27-JAN-1999  
 LOCUS qn194e01.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:1854648 3' similar to gb:J03048 HEMOPEXIN PRECURSOR (HUMAN)  
 ACCESSION A1283497  
 VERSION A1283497.1 GI:3921730  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 494)  
 NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-@email.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 3834 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 380.  
 FEATURES  
 source  
 1.494  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1854648"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /note="Organ: pooled; Vector: pT7TD-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19, testis NHT, and B-cell  
 NCI CGAP GC91) were mixed and as circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 119 a 126 c 145 g 104 t  
 ORIGIN

Query Match 10.9%; Score 406.6; DB 9; Length 494;  
Best Local Similarity 99.0%; Pred. No. 8.7e-95;  
Matches 409; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1479	CAGGACGGCGGCTGTGTGGCTGGACCTGAAGTCAGAGAGCCCAAGCCACGTGACAGAGC	1538
DB	413	CAGGACGGCGGCTGTGTGGCTGGACCTGAAGTCAGAGAGCCCAAGCCACGTGACAGAGC	354
QY	1539	TTCCCTGGGCCCCGTGAAGAGTGAACGAGCCCTTGTGTATGAAAGTCCCTTGCCCTTA	1598
DB	353	TTCCCTGGGCCCCGTGAAGAGTGAACGAGCCCTTGTGTATGAAAGTCCCTTGCCCTTA	294
QY	1599	ACTCATGTTCCGCCAATGATCCCGGCTTGTACTCATCCATGGTCCCAATTGTACTGCT	1658
DB	293	ACTCATGTTCCGCCAATGATCCCGGCTTGTACTCATCCATGGTCCCAATTGTACTGCT	234
QY	1659	ACAGTATGTGAGAACTGAAATGACAGCCCAAGGCCCTTCGCAACCCCAAGATGTACCA	1718
DB	233	CCAGTATGTGAGAACTGAAATGACAGCCCAAGGCCCTTCGCAACCCCAAGATGTACCA	174
QY	1719	GTCTCTGGGCTGCACTCACTGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTC	1778
DB	173	GTCTCTGGGCTGCACTCACTGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTC	114
QY	1779	CTAGTTCCTCATATAAAGACAGATTGCTTCTGGCTTCTCACTGAAGGGGCTTCTGACA	1838
DB	113	CTAGTTCCTCATATAAAGACAGATTGCTTCTGGCTTCTCACTGAAGGGGCTTCTGACA	54
QY	1839	TGAGTCTGGGCTGGCCCACTCCCAAGTTTCTCATATAAAGACAGATTGCT	1891
DB	53	TGAGTCTGGGCTGGCCCACTCCCAAGTTTCTCATATAAAGACAGATTGCT	1

Search completed: December 16, 2003, 05:34:05  
Job time : 4682.61 secs

THIS PAGE BLANK (USPTO)

FEATURES		BASE COUNT		ORIGIN	
Source		129 a	153 c	140 g	129 t
<p>Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience &amp; Biotechnology 52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 4 row: C column: 01 High quality sequence stop: 551.</p>					
<p>Location/Qualifiers</p> <p>1..551</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="L17N670205-4-C01"</p> <p>/sex="F"</p> <p>/lab_host="Top10P"</p> <p>/clone_lib="L17N670205"</p> <p>/note="Organ: Liver; Vector: pUT73-Pac; Site_1: EcoRI; Site_2: NotI; The library was contributed by the Soares Laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."</p>					
<p>Query Match 11.4%; Score 425.8; DB 14; Length 551; Best Local Similarity 99.5%; Pred. No. 8.7e-100; Matches 427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p>					
QY	1479 CAGAGCGCGGCTGTGTGCTGTGACCTGGAAGTCAGAGCCCAAGCCACGTGAGACAGC	1538			
Db	103 CAGACCGCGGCTGTGTGCTGTGACCTGGAAGTCAGAGCCCAAGCCACGTGAGACAGC	162			
QY	1539 TTCCTGGCCCCCAGAGAGGAGAGCGAGCCTGTGTATGAGAAAGTCCCTGGCCCTA	1598			
Db	163 TTCCTGGCCCCCAGAGAGGAGAGCGAGCCTGTGTATGAGAAAGTCCCTGGCCCTA	222			
QY	1599 ACTCATGTTCCGCGCAATGATGTCGCCGCTTGTAACCTCATGATGTCCTCAATTTGTACTGCT	1658			
Db	223 ACTCATGTTCCGCGCAATGATGTCGCCGCTTGTAACCTCATGATGTCCTCAATTTGTACTGCT	282			
QY	1659 ACAAGTATGTGAGAACTGAATGACGACCAAGCCCTTCGCAACCCCAAGATGTACCA	1718			
Db	283 ACAAGTATGTGAGAACTGAATGACGACCAAGCCCTTCGCAACCCCAAGATGTACCA	342			
QY	1719 GTCTCCGCGGCTGCACTCAGTAGGGGGCTTCTGACATGATGTGCGCTTGAGCCCACTC	1778			
Db	343 GTCTCCGCGGCTGCACTCAGTAGGGGGCTTCTGACATGATGTGCGCTTGAGCCCACTC	402			
QY	1779 CTAGTTCCTCAATTAAGACAGATTTGCTTCTTGCTTCTGACTGAGGGGCTTCTGACA	1838			
Db	403 CTAGTTCCTCAATTAAGACAGATTTGCTTCTTGCTTCTGACTGAGGGGCTTCTGACA	462			
QY	1839 TGAAGTCGCGCTGGGCCCACTCCCAAGTTTCTCAATTAAGACAGATTTGCTTCTGAC	1898			
Db	463 TGAAGTCGCGCTGGGCCCACTCCCAAGTTTCTCAATTAAGACAGATTTGCTTCTGAC	522			
QY	1899 TTGAATCAA 1907				
Db	523 TTGAATAAA 531				
<p>RESULT 5 AV645403/c AV645403 704 bp mRNA linear EST 15-JAN-2002</p> <p>DEFINITION AV645403 GLA Homo sapiens cDNA clone GLAAH08 3', mRNA sequence.</p> <p>ACCESSION AV645403</p> <p>VERSION AV645403.1 GI:9866417</p> <p>KEYWORDS EST.</p> <p>SOURCE Homo sapiens (human)</p> <p>ORGANISM Homo sapiens</p>					

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE	21625106
PUBMED	11752456
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel.: 86-21-50801919 (ex.45) Tel.: 86-21-50801922 Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.
FEATURES	Location/Qualifiers
SOURCE	1. 704 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="GLA01A08" /tissue_type="corresponding non cancerous liver tissue" /dev_stage="adult" /lab_host="SOLR" /clone_id="GLA" /note="Vector: pbluescript sk(-); Site_1: EcorI; Site_2: XhoI"
BASE COUNT	167 a 181 c 197 g 155 t 4 others
ORIGIN	
Query Match	11.4%; Score 425.8; DB 9; Length 704;
Best Local Similarity	99.5%; Pred. No. 9.9e-100;
Matches 427; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
1479	CAGAGACGGCGCTGTGGTGCCTGAGCTGAAGTCAGAGCCCAAGCCAGTGGACAGAGC 1538
Db	433 CAGAGACGGCGCTGTGGTGCCTGAGCTGAAGTCAGAGCCCAAGCCAGTGGACAGAGC 374
QY	1539 TTCCTTGGCCCATGAGAAAGGTAGACGAGGCTTGTGTATGAAAAAGTCCCTTGGCCCTTA 1598
Db	373 TTCCTTGGCCCATGAGAAAGGTAGACGAGGCTTGTGTATGAAAAAGTCCCTTGGCCCTTA 314
QY	1599 ACTGATGTCGGCGCAATGATGCCGGGCTTGTAACCTCATGCATGATGATGCCAATTGTAAGTCT 1658
Db	313 ACTGATGTCGGCGCAATGATGCCGGGCTTGTAACCTCATGCATGATGATGCCAATTGTAAGTCT 254
QY	1659 ACAAGTATGTCGAGAACTGAAATGCACGACCAAGGCCCTTCGCAACCCCAAGATGTGACCA 1718
Db	253 ACAAGTATGTCGAGAACTGAAATGCACGACCAAGGCCCTTCGCAACCCCAAGATGTGACCA 194
QY	1719 GTCTTCCTGGGGCTGACATCTGAGAGGGGCTTCTGACATGAATCTGGGCTGGCCCCCACTTC 1778
Db	193 GTCTTCCTGGGGCTGACATCTGAGAGGGGCTTCTGACATGAATCTGGGCTGGCCCCCACTTC 134
QY	1779 CTAGTTCCCTCATATTAAGAAGAGATTCCTTTGCTTCTTCATCATGAGAGGGGCTTCTTGACA 1838
Db	133 CTAGTTCCCTCATATTAAGAAGATTCCTTTGCTTCTTCATCATGAGAGGGGCTTCTTGACA 74
QY	1839 TGAATCTGGGCTGGCCCCCACTCCCAAGTTTCTCATATTAAGAAGAGATTCCTTCTTAC 1898
Db	73 TGAATCTGGGCTGGCCCCCACTCCCAAGTTTCTCATATTAAGAAGAGATTCCTTCTTAC 14
QY	1899 TTGAATCAA 1907
Db	13 TTGAAAAA 5

1233 ACCCTCATGGGATTTATCTGACTGTGTGGATCGGCTTTATCTGCGGCTTCT 1292  
|||  
518 ACCCTCATGGGATTTATCTGACTGTGTGGATCGGCTTTATCTGCGGCTTCT 577  
|||  
QY 1293 CGGCTCATATCATGCGAGGTGAAGGCT 1321  
|||  
Db 578 CGGCTCATATCATGCGAGGTGAAGGCT 606  
|||  
RESULT 3  
BX325177/c 992 bp mRNA linear EST 02-MAY-2003  
LOCUS BX325177 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0D010YE24 3-PRIME, mRNA sequence.  
BX325177  
ACCESSION BX325177.1 GI:30340447  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 992)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0A010BC12NP1&cluster=5958.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0A010BC12NP1.  
FEATURES  
Source  
1..992  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D010YE24"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT  
10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 228 a 270 c 293 g 199 t 2 others  
ORIGIN  
Query Match 11.5%; Score 429.8; Db 13; Length 992;  
Best Local Similarity 73.3%; Pred. No. 1.1e-100; Indels 253; Gaps 2;  
Matches 704; Conservative 1; Mismatches 3;  
QY 928 TCACAGGAGGACCCACTACTGCGCTCTGACACACGCGGATGCTGACATAGTGGCC 987  
|||  
Db 708 TTCAGTGGAGCCACACTACTGCGCTCTGACACACGCGGATGCTGACATAGTGGCC 649  
|||  
QY 988 ATTGCTCATCATGTCGCCCAAGGCTCTTCAGACAGTGCATGCTGCTTTCTGGGAAGA 1047  
|||  
Db 648 ATTGCTCATCATGTCGCCCAAGGCTCTTCAGACAGTGCATGCTGCTTTCTGGGAAGA 589  
|||  
QY 1048 AAACCTCATCTGCTCCAGGTGTATTTGGGGAGAGAGCTTGAGTGAAGACTGGACAAG 1107  
|||  
Db 588 AAACCTCATCTG----- 576  
|||  
QY 1108 CATATCAACTCTGTATTATATACATCTTTGCTCCAGGAGCCAGGTATATGTCT 1167  
|||

Db 575 -----TCCAGGGACCCAGTATATGCT 552  
|||  
QY 1168 TCCTGACAAAGAGAGCTATACCTAGTAAAGGCTTATCCGAAGGCTGGAAAGAG 1227  
|||  
Db 551 TCCTGACAAAGAGAGCTATACCTAGTAAAGGCTTATCCGAAGGCTGGAAAGAG 492  
|||  
QY 1228 TCGGAGACCCCTCATGGGATTTATCTGACTGTGTGGATCGGCTTTATCTGCGGCT 1287  
|||  
Db 491 TCGGAGACCCCTCATGGGATTTATCTGACTGTGTGGATCGGCTTTATCTGCGGCT 432  
|||  
QY 1288 CTTCGAGCTCATATCATATGAGGAGGTGAGGGGCTTCGAGTGTAGAGGAGCTTGT 1347  
|||  
Db 431 CTTCGAGCTCATATCATATG----- 411  
|||  
QY 1348 CTGCTACCTGTCTGTGGCATATATCCCAACAGGGGATGAGAGGCTAGTCAGATCC 1407  
|||  
Db 410 ----- 411  
|||  
QY 1408 CCAGGAGATGAGAGGCTAGTCAAGATCCCATGACATGAGAGCCATGCTATGTTGG 1467  
|||  
Db 410 ----- 411  
|||  
QY 1468 TGCCTTCTCCAGAGAGCGGCTGTGTGTGCTGACCTGAAGTCAAGAGCCAGCCAC 1527  
|||  
Db 410 -----CAGACCGGCGGCTGTGTGTGCTGACCTGAAGTCAAGAGCCAGCCAC 362  
|||  
QY 1528 GTGAGACAGCTTCTTGGCCCCCATGAGAGAGTGAACGAGCTTGTATGAGAAAGTC 1587  
|||  
Db 361 GTGAGACAGCTTCTTGGCCCCCATGAGAGAGTGAACGAGCTTGTATGAGAAAGTC 302  
|||  
QY 1588 CTTGGACCTTAATCTCATGTTCGCGCAATGATGCTCCGCTTGTACTCATATGATGCCAA 1647  
|||  
Db 301 CTTGGACCTTAATCTCATGTTCGCGCAATGATGCTCCGCTTGTACTCATATGATGCCAA 242  
|||  
QY 1648 TTTGATCTGCTACATGATGTGAGAGAACTGAATGAGGAGCCCTTCCGCAACCCA 1707  
|||  
Db 241 TTTGATCTGCTACATGATGTGAGAGAACTGAATGAGGAGCCCTTCCGCAACCCA 182  
|||  
QY 1708 GAATGTGACAGTCTCTGCGGCTGACATCTGAGGGGCTTCTGACATGATGTGGCT 1767  
|||  
Db 181 GAATGTGACAGTCTCTGCGGCTGACATCTGAGGGGCTTCTGACATGATGTGGCT 122  
|||  
QY 1768 GGCCCCACCTCTAGTCTCTCATATTAAGACAGATGCTTCTGCTCTGACATGAGG 1827  
|||  
Db 121 GGCCCCACCTCTAGTCTCTCATATTAAGACAGATGCTTCTGCTCTGACATGAGG 62  
|||  
QY 1828 GCTTCTGACATGATGTGCGCTGCGCCACCTCCCACTTCTCATATTAAGACAGAT 1887  
|||  
Db 61 GCTTCTGACATGATGTGCGCTGCGCCACCTCCCACTTCTCATATTAAGACAGAT 2  
|||  
QY 1888 T 1888  
|||  
Db 1 T 1  
|||  
RESULT 4  
CB154152 551 bp mRNA linear EST 29-JAN-2003  
LOCUS CB154152  
DEFINITION K-BST0211925 L17N670205 Homo sapiens cDNA clone L17N670205-4-C01  
5', mRNA sequence.  
CB154152  
ACCESSION CB154152.1 GI:28139149  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 551)  
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
Kim, Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished



Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

## FEATURES

## source

Location/Qualifiers

1. .632  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="GLCDM09"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_idb="GIC"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 225 a 97 c 164 g 145 t 1 others  
ORIGIN

Query Match 15.6%; Score 583.4; DB 9; Length 632;  
Best Local Similarity 98.7%; Pred. No. 5.7e-141;

Matches 598; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 34 CAGATTGGAAGTCATCAGCATATATGATGCAATTTAAAGCCCTGAGCTAGGTGAGATTAC 93  
DB 1 CAGATTGGAAGTCATCAGCATATATGATGCAATTTAAAGCCCTGAGCTAGGTGAGATTAC 60  
QY 94 CAAAGAAAGTGAAGTGAAGAGAAAGAGAGCCCAAGTGAAGGAGATTCCAAATTATTA 153  
DB 61 CAAAGAAAGTGAAGTGAAGAGAAAGAGAGCCCAAGTGAAGGAGATTCCAAATTATTA 120  
QY 154 GATATCAGTGTGAAGAAAGAGTGTCAAAAAGATTAAGAGAAATCTGGAGAGAGTACG 213  
DB 121 GATATCAGTGTGAAGAAAGAGTGTCAAAAAGATTAAGAGAAATCTGGAGAGAGTACG 180  
QY 214 TGTCAAGAAAGCCAAAGTTCCAAAAAGACATTTAAAGAGAAAGATGAGCAGTC 273  
DB 181 TGTCAAGAAAGCCAAAGTTCCAAAAAGACATTTAAAGAGAAAGATGAGCAGTC 240  
QY 274 CAGTGTCTCTGAAGAGTGAAGTGAAGAGAAATGACATGAGATTTCGCAA 333  
DB 241 CAGTGTCTCTGAAGAGTGAAGTGAAGAGAAATGACATGAGATTTCGCAA 300  
QY 334 TTGAGAAATCTAGCAACTGATTAAGAAACAATTTCAATGTTGAGGAGAAACGAAGTGT 393  
DB 301 TTGAGAAATCTAGCAACTGATTAAGAAACAATTTCAATGTTGAGGAGAAACGAAGTGT 360  
QY 394 AATTGAAGAGATTGAGAAAAAGACAAATGGAGGCTTAGATTAATCTTAAATTAAGTTG 453  
DB 361 AATTGAAGAGATTGAGAAAAAGACAAATGGAGGCTTAGATTAATCTTAAATTAAGTTG 420  
QY 454 TTGTGAAAAAGAGAAAGAAAGAGGGGTGTAGCCAGCTACTCTCTCACTTTCCACC 513  
DB 421 TTGTGAAAAAGAGAAAGAAAGAGGGGTGTAGCCAGCTACTCTCTCACTTTCCACC 480  
QY 514 ACCTCATAGGAGAGAGCTGGAGAACACAGCCAGAGTGAAGCAATTCAGTGAAGTGTGC 573  
DB 481 ACCTCATAGGAGAGAGCTGGAGAACACAGCCAGAGTGAAGCAATTCAGTGAAGTGTGC 539  
QY 574 TTCTTTTAAAGTTCTGAGACATGTATTTCTATATCTAATACCGCATCTCTGTACATGA 633  
DB 540 TTCTTTTAAAGTTCTGAGACATGTATTTCTATATCTAATACCGCATCTCTGTACATGA 599  
QY 634 CACCTG 639  
DB 600 CACCTG 605

RESULT 2  
AM965111 672 bp mRNA linear EST 01-JUN-2000  
LOCUS  
DEFINITION ESTJ77184 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
VERSION AM965111.1 GI:8154947

KEYWORDS  
SOURCE  
ORGANISM

EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 672)  
Hegde, P., Qi, R., Abernathy, K., Dharrap, S., Gaepard, R., Gay, C., Holt  
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and

## AUTHORS

Quackenbush, J.

## TITLE

Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray

## JOURNAL

Unpublished  
Contact: John Quackenbush

## COMMENT

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johng@cigr.org  
Plate: 210

## FEATURES

## source

Location/Qualifiers

1. .672

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

/clone\_idb="MAGE resequences, MAGI"  
/note="Vector: pBluescriptSkm"

BASE COUNT 128 a 205 c 167 g 172 t  
ORIGIN

Query Match 13.6%; Score 506.6; DB 10; Length 672;  
Best Local Similarity 87.4%; Pred. No. 7.2e-121;

Matches 602; Conservative 0; Mismatches 4; Indels 83; Gaps 1;

QY 633 ACACTGAATCTCTTGAAGAGTGCAGCCCAAGCCCAAGTGTGGCTTACCTGAAGAACT 692  
DB 1 ACACTGAATCTCTTGAAGAGTGCAGCCCAAGCCCAAGTGTGGCTTACCTGAAGAACT 60  
QY 693 TAGCACTGTTTCCACACTTGTCTTTTCAAGGCACTGCTGATTCCAGTTCCAGCCA 752  
DB 61 TAGCACTGTTTCCACACTTGTCTTTTCAAGGCACTGCTGATTCCAGTTCCAGCCA 120  
QY 753 GGGCAGAGTGGCCCAATGCTGAGCAAGCTTGTCTAATTTCTCCTTCACTGAGGCT 812  
DB 121 GGGCAGAGTGGCCCAATGCTGAGCAAGCTTGTCTAATTTCTCCTTCACTGAGGCT 180  
QY 813 CTTCATCTTGGCTCTGATGATTCCTCCTCTCATAGCACTATTTCTGATTCATCA 872  
DB 181 CTTCATCTTGGCTCTGATGATTCCTCCTCTCATAGCACTATTTCTGATTCATCA 240  
QY 873 CTAGCTCTTCTTGTGCTTGGCTTGTGCGAGCGGCTTAGAGCACTATGATTCAC 932  
DB 241 CTAGCTCTTCTTGTGCTTGGCTTGTGCGAGCGGCTTAGAGCACTATGATTCAC 300  
QY 933 AGGGAACCACTACTAGGCTGTGAGCAACAGCGGGATGCTGAGCATAGCTGGCCATTGC 992  
DB 301 AGGGAACCACTACTAGGCTGTGAGCAACAGCGGGATGCTGAGCATAGCTGGCCATTGC 360  
QY 993 TCATCAGTGGCCCAAGGCTCTTCAAGCAGTGTGCTTTTCTGGAAGAAAAACT 1052  
DB 361 TCATCAGTGGCCCAAGGCTCTTCAAGCAGTGTGCTTTTCTGGAAGAAAAACT 420  
QY 1053 CTATCTGTGTCAGAGTGTATTTGGGGAGAGGCTTAGAGTAGAGCTGGGACAAGATAT 1112  
DB 421 CTATCTGTG----- 428  
QY 1113 CCAACTCTGATTTATTAACATCTTGTCTCAGAGGCAAGGATATATGCTTCTG 1172  
DB 429 -----TCAGAGGCAAGGATATATGCTTCTG 457  
QY 1173 ACAAGAGAGGCTATACCTAGTAGCGGTTATCCAGACGAGCTGAGAGAGAGTCCGG 1232  
DB 458 ACAAGAGAGGCTATACCTAGTAGCGGTTATCCAGACGAGCTGAGAGAGAGTCCGG 517



Job time : 780.679 BECS

Qy	1136	CTTGTGCTCCAGGGACCCGAGATATAGTCTTCTGCAAGAGGAGCTTACCCAGT	1159
Db	93	CTACTCATCAGGGTACCCAGGTAATATCTTCTGCAAGGGGAGGCTTCACTCTGCT	152
Qy	1196	AAGCGGTTATCCGAAGCGGCTGGAGAGGAAGTGGGACCCCTCATGGGATTAATCTGGA	1255
Db	153	AAAGATTAATCCAAAGCAGCTGGAGGAAGAAATTGGAGCCCTGATGGGGTCTGGCTTCA	212
Qy	1256	CTCTGTGGAATCGGAGCTTTATCTGCGCTTCTTCTGGGCTCCATATCATGGCAGGTGA	1315
Db	213	TTCTGTGATACAGGCTTTACCTGCTCTGGATCTTCTCAGCTCTCATCATGGCAGGCCA	272
Qy	1316	GGGGCT	1321
Db	273	GAACTT	278

## RESULT 15

US-09-960-352-5368  
 ; Sequence 5368, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 5368  
 ; LENGTH: 374  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 23-LIB34-009-Q1-E1-F11  
 US-09-960-352-5368

Query Match	3.2;	Score	118.2;	DB	10;	Length	374;
Best Local Similarity	64.1;	Pred	NO. 3.1e-27;				
Matches	243;	Conservative	0;	Mismatches	53;	Indels	83;
						Gaps	1;

Qy	927	TTCCACAGGAGCCACATCTAGTGGCGCTGTGACACACAGCGGGAGTGGCTGGATAGCTGGCC	986
Db	79	TTTCAGTGAAGAACATCTAGCTGGCGCTCTTGGACAGACAGCCGGGAGCGGTGGACACAGCTGGCC	138
Qy	987	CATTGCTCATCAGTGGGCCCAAGGGTCCTTCAGCAGTGGATGTGCGCTTTTCTCTGGAGAGA	1048
Db	139	CATTGAGCATCTGTGGCCCCCAAGTCCCTCAACAGTGAATGTGCTTTCTCTGGAGATA	198
Qy	1047	AAAACTATCTGGTCCAGGTGTGTATTGGGGAGAGGCTTGAGGTAGAGACTGGGACAA	1106
Db	199	AAAGCTTACC-----	209
Qy	1107	GCATATCCAACTGTATTTATTATACATCTTTGCTCCAGGGACCCAGGTATATGTC	1166
Db	210	-----TCATCCAGGGATACCAGGTATATATTC	235
Qy	1167	TTCTCGACAAAGGAGGCTATATCCCTAGTAAGCGGTTATCCGAACGGCTGGAGAGGAA	1228
Db	236	TTCTCGACAAAGGCGAGGCTACACTCTCGTAAAGATTTCCAAACAGCTGGAGAGGAA	295
Qy	1227	GTCGGGACCCCTCATGGGATTAATCTTGAACCTGTGGAATGCGGCTTTATCTGCCCTG3GG	1286
Db	296	TTTGGGAGCCCTGATGAGGGGTCTGCGCTTCAATCTGTGGAATGAGAGCTTAACTGTGCTGGA	355
Qy	1287	TCTTCTGGGCTCCATATCA	1305
Db	356	TCCTTTCAGCTCTTACATCA	374

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(488)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30754

Query Match 4.0%; Score 149.4; DB 11; Length 488;  
Best Local Similarity 73.4%; Pred. No. 2.3e-37;  
Matches 246; Conservative 0; Mismatches 6; Indels 83; Gaps 1;

QY 928 TCACAGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 987  
DB 237 TTCAGTGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 236  
QY 988 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 1047  
DB 297 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 356  
QY 1048 AAATCTATCTGGTCCAGGTGTGTATATGGGAGAGGCTTGAAGTGAAGACTGGACAAG 1107  
DB 357 AAATCTATCTGGTCCAGGTGTGTATATGGGAGAGGCTTGAAGTGAAGACTGGACAAG 371  
QY 1108 CATATCCAACTCTGTATTATTATTAACATCTTTGTCTCAGGACCCAGATATATGCTT 1167  
DB 372 -----CAGGGACCCAGATATATGCTT 393  
QY 1168 TCCTGCAAGAGGAGGCTATACCTTAGTAGCGGTTATCCGAAGCGGCTGGAAGAGAG 1227  
DB 394 TCCTGCAAGAGGAGGCTATACCTTAGTAGCGGTTATCCGAAGCGGCTGGAAGAGAG 453  
QY 1228 TCGGAGCCCTCATGGGATATATCTGACTCTGTG 1262  
DB 454 ACGGAGCCCTCATGGGATATATCTGACTCTGTG 488

RESULT 12  
US-09-918-995-32181  
Sequence 32181, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32181  
LENGTH: 473  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(473)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-32181

Query Match 3.6%; Score 134.2; DB 11; Length 473;  
Best Local Similarity 97.8%; Pred. No. 2.1e-32;  
Matches 136; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 928 TCACAGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 987  
DB 323 TTCAGTGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 382  
QY 988 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 1047  
DB 383 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 442

QY 1048 AAATCTATCTGGTCCAGG 1066  
DB 443 AAATCTATCTGGTCCAGG 461

RESULT 13  
US-09-918-995-32820  
Sequence 32820, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32820  
LENGTH: 491  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(491)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-32820

Query Match 3.6%; Score 134.2; DB 11; Length 491;  
Best Local Similarity 97.8%; Pred. No. 2.2e-32;  
Matches 136; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 928 TCACAGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 987  
DB 272 TTCAGTGGAGGACCACTACTGCGCTTGAGACACCGCGGATGCTGGCATAGTGGCCC 331  
QY 988 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 1047  
DB 332 ATTGCTCATAGTGGCGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTGGGAAGA 391  
QY 1048 AAATCTATCTGGTCCAGG 1066  
DB 392 AAATCTATCTGGTCCAGG 410

RESULT 14  
US-09-960-352-10323  
Sequence 10323, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalegan, Nagappan  
TITLE OF INVENTION: NOVEL NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 10323  
LENGTH: 420  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 44-LIB34-041-Q1-B1-C8  
US-09-960-352-10323

Query Match 3.3%; Score 123.6; DB 10; Length 420;  
Best Local Similarity 79.0%; Pred. No. 5.8e-29;  
Matches 147; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 28-LIB34-031-Q1-E1-G7  
US-09-960-352-6580

Query Match 5.3%; Score 197.4; DB 10; Length 400;  
Best Local Similarity 75.8%; Pred. No. 3.9e-53;  
Matches 257; Conservative 0; Mismatches 81; Indels 1; Gaps 1;

QY 1479 CAGAGCGCGGCTGTGTGCTGCTGACCTTGAAGTCAGAGCCCAAGCCACCTGACAGAGC 1538  
DB 366 CAGGCCAGAGGCTGTCAAGCTGTGACCTTGAAGTGAAGCTCAAGCCACCTGACAGAGA 307  
QY 1539 TTCCTTGGCCCTGAGAGGATGAGAGGAGCTGTGTATGGAAGATCCCTTGGCCCTTA 1598  
DB 306 TTCCTTGGCTCCATACAGAAAGTGAATGGAGGCTGTGTATGAGAAATATCTGGCCCC 247  
QY 1599 ACTCATGTTCGCGCAATGATGCCGCTTGTATCACTCATCATGTGTCCCAATTTGACTGCT 1658  
DB 246 ACTCGATGTGCAATGATGTCTGGGCTTGTACCTCGTCAAGGCCCAATCTGTACTGCT 187  
QY 1659 ACAGTATGTGAGAAATGTAATGACAGCCAGGCCCTTCCGCAACCAAGATGTGACCA 1718  
DB 186 ACAAGATGTGAGAGATTTAGACAGACCAAGACCTTCCGCAAGGCCCAAGAGATGAACA 127  
QY 1719 GTCTCCGGGCTGCACTCACTGAGGGGCTTTCGACATGATGTGGCTGGCCCACTC 1778  
DB 126 GCCTTCGGGCTGCTGCCCTCCCAACAGCACTCTGATCGAATTTGGCTCAAGATCTCTC 67  
QY 1779 CTAG-TTCTCATATTAAGACAGATTTGCTTCTGCTT 1816  
DB 66 CCAATTTCATCATATTAAGCCAGAGTCTTCTTCACTT 28

RESULT 9  
US-09-960-352-11316  
Sequence 11316, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengding  
APPLICANT: Byatt, John C.  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960.352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 11316  
LENGTH: 387  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 48-LIB34-079-Q1-E1-D8  
US-09-960-352-11316

Query Match 4.8%; Score 180.6; DB 10; Length 387;  
Best Local Similarity 78.8%; Pred. No. 1.2e-47;  
Matches 216; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1479 CAGAGCGCGGCTGTGTGCTGCTGACCTTGAAGTCAGAGCCCAAGCCACCTGACAGAGC 1538  
DB 111 CAGGCCAGAGGCTGTGAGGCTGGAAGCTTGAAGAGCTCAAGCCACCTGACAGAGC 170  
QY 1539 TTCCTTGGCCCTGAGAGGATGAGAGGAGCTTGTATGGAAGATCCCTTGGCCCTTA 1598  
DB 171 TTCCTTGGCTCCATACAGAAAGTGAATGGAGGCTGTGTATGAGAAATATCTGGCCCC 230  
QY 1599 ACTCATGTTCGCGCAATGATGCCGCTTGTATCACTCATCATGTGTCCCAATTTGACTGCT 1658  
DB 231 ACTCGATGTGCAATGATGTCTGGGCTTGTACCTCGTCAAGGCCCAATCTGTACTGCT 290  
QY 1659 ACAGTATGTGAGAAATGTAATGACAGCCAGGCCCTTCCGCAACCAAGATGTGACCA 1718

DB 291 ACAAGATGTGAGAGATTTAGAGCAAGACCAAGGACCTTCCCAAGGCCCAAGAGATGAACA 350  
QY 1719 GTCTCCGGGCTGTCACTCACTGAGGGGCTTCTGTA 1753  
DB 351 GCCTCCTGGGCTGGCTCCCAACAGCACTCTGTA 385

RESULT 10  
US-09-960-352-7040  
Sequence 7040, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengding  
APPLICANT: Byatt, John C.  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960.352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 7040  
LENGTH: 408  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 30-LIB34-026-Q1-E1-H5  
US-09-960-352-7040

Query Match 4.5%; Score 167; DB 10; Length 408;  
Best Local Similarity 79.8%; Pred. No. 3.5e-43;  
Matches 197; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1479 CAGAGCGCGGCTGTGTGCTGCTGACCTTGAAGTCAGAGCCCAAGCCACCTGACAGAGC 1538  
DB 162 CAGGCCAGAGGCTGTGAGGCTGGAAGCTTGAAGAGCTCAAGCCACCTGACAGAGC 221  
QY 1539 TTCCTTGGCCCTGAGAGGATGAGAGGAGCTTGTATGGAAGATCCCTTGGCCCTTA 1598  
DB 222 TTCCTTGGCTCCATACAGAAAGTGAATGGAGGCTGTGTATGAGAAATATCTGGCCCC 281  
QY 1599 ACTCATGTTCGCGCAATGATGCCGCTTGTATCACTCATCATGTGTCCCAATTTGACTGCT 1658  
DB 282 ACTCGATGTGCAATGATGTCTGGGCTTGTACCTCGTCAAGGCCCAATCTGTACTGCT 341  
QY 1659 ACAGTATGTGAGAAATGTAATGACAGCCAGGCCCTTCCGCAACCCCAAGATGTGACCA 1718  
DB 342 ACAAGATGTGAGAGATTTAGAGCAAGACCAAGGACCTTCCCAAGGCCCAAGAGATGAACA 401  
QY 1719 GTCTCCT 1725  
DB 402 GCCTCCT 408

RESULT 11  
US-09-918-995-30754  
Sequence 30754, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyeeg, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918.995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235.076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: PaeSeq for Windows Version 3.0  
SEQ ID NO 30754  
LENGTH: 488  
TYPE: DNA

Db 60 TGAGTCG--CTGNCACCTCCCGAGTTTCATATAAGACAGATTGCANTTCAC 3  
 Qy 1899 TT 1900  
 Db 2 TT 1

# RESULT 6

US-10-316-253-39  
 ; Sequence 39, Application US/10316253  
 ; Publication No. US20030162706A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Procter & Gamble Company  
 ; APPLICANT: Peters, Kevin  
 ; APPLICANT: Thompson, Larry  
 ; APPLICANT: Wang, Feng  
 ; APPLICANT: Greis, Kenneth  
 ; TITLE OF INVENTION: Angiogenesis Modulating Proteins  
 ; FILE REFERENCE: 8865M  
 ; CURRENT APPLICATION NUMBER: US/10/316,253  
 ; PRIOR FILING DATE: 2002-12-10  
 ; PRIOR APPLICATION NUMBER: US 60/355,295  
 ; PRIOR FILING DATE: 2002-02-08  
 ; NUMBER OF SEQ ID NOS: 308  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 39  
 ; LENGTH: 1516  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (55)..(1437)  
 ; OTHER INFORMATION:  
 ; US-10-316-253-39

Query Match 5.3%; Score 198.6; DB 13; Length 1516;  
 Best Local Similarity 75.1%; Pred. No. 4.2e-53;  
 Matches 262; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

Qy 1471 CTCTCCCGAGAGCGGCGGTGTGTGCTGACCTGAATCGAGAGCCCAAGCCAGTG 1530  
 Db 1167 CGTCAATCATGAGAGCGGCGGCTTTGTGTGCTGACCTGAATCGAGAGCCCAAGCCAGTG 1226  
 Qy 1531 GACAGAGCTTCTTGCGCCCATGAGAGGTAGACGAGCCTTGATGAGAAAGTCCCT 1590  
 Db 1227 GCGAGAGCTTCTTGCGCCCATGAGAGGTAGATGAGTGGCCCTGTGTTGAGAAAGTCCCT 1286  
 Qy 1591 TGGCCCTAATCATGTTCCGCAATGAGTCCGCGCTGTATCATCATGATGTCCTCAATT 1650  
 Db 1287 TGGTCCCTAATCATGTTCCGCAATGAGTCCGCAATGTTCTTATCATGAGCCCAATT 1346  
 Qy 1651 GTACTGCTACAGTGTGAGAAACTGAATGACGAGCCCTTCCGCAATGAGAA 1710  
 Db 1347 ATACTCTATGAGATGATGAGAAACTGAATGACGAGCCCAATGCTCTGAGCCCAAGAA 1406  
 Qy 1711 TGTGACAGTCTCTGCGCTGCACTGAGGGGCC--TTCTGACATGAGTGGCCT 1767  
 Db 1407 AGTGAACAGATCTTGCGCTGCACTGATTAAGAGCCCTGATGAGAAATTAGCCAGCCA 1466  
 Qy 1768 GGGCCCACTCTAGTCTCTCATATTAAGACAGATGCTTTCGCTT 1816  
 Db 1467 CCCCACTCTCTTCATTCATTAATAAACAGATGTTTCTTCAT 1515

RESULT 7  
 US-10-175-523-84  
 ; Sequence 84, Application US/10175523  
 ; Publication No. US20030096264A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brockman, Jeffrey  
 ; APPLICANT: Evans, David  
 ; APPLICANT: Hook, Derek  
 ; APPLICANT: Klimczak, Leszek

APPLICANT: Laeng, Pascal  
 APPLICANT: Palfreyman, Michael  
 APPLICANT: Rajan, Prithi  
 TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)  
 FILE REFERENCE: 3235/10795-US3  
 CURRENT APPLICATION NUMBER: US/10/175,523  
 PRIOR FILING DATE: 2002-06-18  
 PRIOR APPLICATION NUMBER: US 60/299,151  
 PRIOR FILING DATE: 2001-06-18  
 PRIOR APPLICATION NUMBER: US 60/317,828  
 PRIOR FILING DATE: 2001-09-07  
 PRIOR APPLICATION NUMBER: US 60/325,150  
 PRIOR FILING DATE: 2001-09-25  
 PRIOR APPLICATION NUMBER: US 60/333,047  
 PRIOR FILING DATE: 2001-11-14  
 PRIOR APPLICATION NUMBER: US 60/349,936  
 PRIOR FILING DATE: 2002-01-18  
 PRIOR APPLICATION NUMBER: US 60/361,834  
 PRIOR FILING DATE: 2002-03-04  
 NUMBER OF SEQ ID NOS: 197  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 84  
 LENGTH: 1516  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 US-10-175-523-84

Query Match 5.3%; Score 198.6; DB 15; Length 1516;  
 Best Local Similarity 75.1%; Pred. No. 4.2e-53;  
 Matches 262; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

Qy 1471 CTCTCCCGAGAGCGGCGGTGTGTGCTGACCTGAATCGAGAGCCCAAGCCAGTG 1530  
 Db 1167 CGTCAATCATGAGAGCGGCGGCTTTGTGTGCTGACCTGAATCGAGAGCCCAAGCCAGTG 1226  
 Qy 1531 GACAGAGCTTCTTGCGCCCATGAGAGGTAGACGAGCCTTGATGAGAAAGTCCCT 1590  
 Db 1227 GCGAGAGCTTCTTGCGCCCATGAGAGGTAGATGAGTGGCCCTGTGTTGAGAAAGTCCCT 1286  
 Qy 1591 TGGCCCTAATCATGTTCCGCAATGAGTCCGCGCTGTATCATCATGATGTCCTCAATT 1650  
 Db 1287 TGGTCCCTAATCATGTTCCGCAATGAGTCCGCAATGTTCTTATCATGAGCCCAATT 1346  
 Qy 1651 GTACTGCTACAGTGTGAGAAACTGAATGACGAGCCCTTCCGCAATGAGAA 1710  
 Db 1347 ATACTCTATGAGATGATGAGAAACTGAATGACGAGCCCAATGCTCTGAGCCCAAGAA 1406  
 Qy 1711 TGTGACAGTCTCTGCGCTGCACTGAGGGGCC--TTCTGACATGAGTGGCCT 1767  
 Db 1407 AGTGAACAGATCTTGCGCTGCACTGATTAAGAGCCCTGATGAGAAATTAGCCAGCCA 1466  
 Qy 1768 GGGCCCACTCTAGTCTCTCATATTAAGACAGATGCTTTCGCTT 1816  
 Db 1467 CCCCACTCTCTTCATTCATTAATAAACAGATGTTTCTTCAT 1515

RESULT 8  
 US-09-960-352-6580/C  
 ; Sequence 6580, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Ningbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; PRIOR FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 6580  
 ; LENGTH: 400



QY 3568 AGTTAGGCTGCTATAGAGATATCTTAAAGTGGGTATCTATACAGCAATAGGAATTTA 3627  
DB 3017 AGTTAGGCTGCTATAGAGATATCTTAAAGTGGGTATCTATACAGCAATAGGAATTTA 3076  
QY 3628 TTGTTTCACATTTCTGAGGCTGGAATAATCCAGATCAAGGCTCCAGAGGTTCAAGTCT 3687  
DB 3077 TTGTTTCACATTTCTGAGGCTGGAATAATCCAGATCAAGGCTCCAGAGGTTCAAGTCT 3136  
QY 3688 GCTGAGGCTGTTTCTGCTTCCGAAGATGCACTTTTGTCTGTCTCA 3737  
DB 3137 GCTGAGGCTGTTTCTGCTTCCGAAGATGCACTTTTGTCTGTCTCA 3186

## RESULT 3

US-10-125-237-19  
Sequence 19, Application US/10125237  
Publication No. US20030022329A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Zhou, Ping  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyan  
APPLICANT: Zhao, Qing A.  
APPLICANT: Xue, Aidong J.  
APPLICANT: Zhang, Jie  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Dmanac, Radoje T.  
TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and  
FILE REFERENCE: 791CIP2ADIV  
CURRENT APPLICATION NUMBER: US/10/125,237  
CURRENT FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: 09/668,317  
PRIOR FILING DATE: 2000-09-22  
PRIOR APPLICATION NUMBER: 09/552,929  
PRIOR FILING DATE: 2000-04-18  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: pc\_fl\_genes Version 2.0  
SEQ ID NO 19  
LENGTH: 1631  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)..(1459)  
US-10-125-237-19

Query Match 11.9%; Score 446; DB 15; Length 1631;  
Best Local Similarity 73.7%; Pred. No. 4,7e-134;  
Matches 722; Conservative 0; Mismatches 5; Indels 253; Gaps 2;  
QY 928 TCCACAGGAGCCCACTACTGCGCTGAGACACAGCCGGAGTGGTGGACCTGAGCC 987  
DB 899 TTCAAGTGGACCCCACTACTGCGCTGAGACACAGCCGGAGTGGTGGACCTGAGCC 958  
QY 988 ATTGCTCATGAGTGGCCCGAGGCTCTTCAGCAGTGGATGCTCTTTCTGGAGAA 1047  
DB 959 ATTGCTCATGAGTGGCCCGAGGCTCTTCAGCAGTGGATGCTCTTTCTGGAGAA 1018  
QY 1048 AAACCTCATGCTGCTCCAGGTGTGTATTTGGGGAGAGGCTTAGTAGACTGGAGAC 1107  
DB 1019 AAACCTCATGCTGCTCCAGGTGTGTATTTGGGGAGAGGCTTAGTAGACTGGAGAC 1031  
QY 1108 CATATCAACTCTGTATTATTAACATCTTTGCTCCAGGGACCCAGATATATGCT 1167  
DB 1032 -----TCAGGGACCCAGATATATGCT 1055  
QY 1168 TCCTGACAAAGGAGGCTATACCTAGTAGACGTTATCCGAAGCGGCTGGAGAGAG 1227  
DB 1056 TCCTGACAAAGGAGGCTATACCTAGTAGACGTTATCCGAAGCGGCTGGAGAGAG 1115

QY 1228 TCGGAGCCCTCATGAGATTTATCTGAGACTCTGTGATGCGGCTTTATCTGCGCTGG 1287  
DB 1116 TCGGAGCCCTCATGAGATTTATCTGAGACTCTGTGATGCGGCTTTATCTGCGCTGG 1175  
QY 1288 CTTCGAGCTCCATATCATGAGAGGTGAGGGGCTTCGAGTCTTAGAGGGACCTTGT 1347  
DB 1176 CTTCGAGCTCCATATCATG----- 1196  
QY 1348 CTGCTACTGTCTGTGAGCATAGATCCCAACAGGGCATAGAGAGGCTTAGATCC 1407  
DB 1197 ----- 1196  
QY 1408 CAGGGCATGAGAGGCTTAGATCCCAACAGGGCATAGAGAGGCTTAGATCC 1467  
DB 1197 ----- 1196  
QY 1468 TGCCTTCTCCCAAGAGCGGCTGTGTGCTGAGACTGAACTCAGAGCCCAAGCCAC 1527  
DB 1197 -----CAGAGCGGCTGTGTGCTGAGACTGAACTCAGAGCCCAAGCCAC 1245  
QY 1528 GTGAGAGAGCTTCTTGGCCCCCATGAGAGAGGTAGACGAGGCTTGTATGAGAAAGTC 1587  
DB 1246 GTGAGAGAGCTTCTTGGCCCCCATGAGAGAGGTAGACGAGGCTTGTATGAGAAAGTC 1305  
QY 1588 CCTTGGCCCTTAATCATATGTTCCGCAATGATGCTCCGCTTGTACTCATATGATCCCA 1647  
DB 1306 CCTTGGCCCTTAATCATATGTTCCGCAATGATGCTCCGCTTGTACTCATATGATCCCA 1365  
QY 1648 TTGTATCTGCTACAGTATGAGAGAACTGATGAGCCCAAGGCTTCCGCAACCCCA 1707  
DB 1366 TTGTATCTGCTACAGTATGAGAGAACTGATGAGCCCAAGGCTTCCGCAACCCCA 1425  
QY 1708 GAATGACACAGTCTTCTGGGCTGACACTGAGAGGCTTCTGACATGAGTCTGCT 1767  
DB 1426 GAATGACACAGTCTTCTGGGCTGACACTGAGAGGCTTCTGACATGAGTCTGCT 1485  
QY 1768 GGGCCCACTCTAGTCTCTCATATTAAGACAGATGCTTCTGCTCTCATGAGG 1827  
DB 1486 GGGCCCACTCTAGTCTCTCATATTAAGACAGATGCTTCTGCTCTCATGAGG 1545  
QY 1828 GCTTCTGACATGAGTGGGCTGAGCCCACTCCCACTTCTCATTAATAAGACAGAT 1887  
DB 1546 GCTTCTGACATGAGTGGGCTGAGCCCACTCCCACTTCTCATTAATAAGACAGAT 1605  
QY 1888 TGTCTTCACTTGAATCAA 1907  
DB 1606 TGTCTTCACTTGAATCAAAAA 1625

## RESULT 4

US-10-105-891-19  
Sequence 19, Application US/10105891  
Publication No. US20030073099A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Zhou, Ping  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyan  
APPLICANT: Zhao, Qing A.  
APPLICANT: Xue, Aidong J.  
APPLICANT: Zhang, Jie  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Dmanac, Radoje T.  
TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and  
FILE REFERENCE: 791CIP2A  
CURRENT APPLICATION NUMBER: US/10/105,891  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER: 09/668,317  
PRIOR FILING DATE: 2000-09-22



Db	928	-----	927
Qy	1408	CCAGGGCATGAGAAAGCCCTAGGTGACAGATCCCATGACATGAAAGCCATGCTATGTTGG	1467
Db	928	-----	927
Qy	1468	TGCTCTTCCCCAGAGACGGGGGCTGTGGTGGCTGACCTGAAGTCAAGAGCCCAAGCCAC	1527
Db	928	-----CAGACAGGGCGGCTGTGGTGGCTGACCTGAAGTCAAGAGCCCAAGCCAC	976
Qy	1528	GTGGAAGAGCTTCTTGGCCCCATGAGAAAGTGAACGAGACCTTGTGTATGAAAAATGC	1587
Db	977	GTGGAAGAGCTTCTTGGCCCCATGAGAAAGTGAACGAGACCTTGTGTATGAAAAATGC	1036
Qy	1588	CCTTGGCCCTTAACCTCATGTTTCCGCCAATGTGTCGGGCTTGTACTCATCTGATGCCCA	1647
Db	1037	CCTTGGCCCTTAACCTCATGTTTCCGCCAATGTGTCGGGCTTGTACTCATCTGATGCCCA	1096
Qy	1648	TTTGTACTGCTACAGTGAATGTGAGAAACGTAAATGACGCCAAAGGCCCTTCGCAACCCCA	1707
Db	1097	TTTGTACTGCTACAGTGAATGTGAGAAACGTAAATGACGCCAAAGGCCCTTCGCAACCCCA	1156
Qy	1708	GAATGTGACAGTCTCTCGGGCTGCACCTCACTGAGGGGCTTCTGACATGAGTCTGGCCT	1767
Db	1157	GAATGTGACAGTCTCTCGGGCTGCACCTCACTGAGGGGCTTCTGACATGAGTCTGGCCT	1216
Qy	1768	GGCCCCACCTCTTGAAGTCTCTCATATATAAAGACAGTTGCTTCTTCTCACTGAGGG	1827
Db	1217	GGCCCCACCTCTTGAAGTCTCTCATATATAAAGACAGTTGCTTCTTCTCACTGAGGG	1276
Qy	1828	GGCTTCTGACATGAGTCTGGGCCCTGGCCCCACCTCCCGAGTTTCTCATATATAAAGACAGT	1887
Db	1277	GGCTTCTGACATGAGTCTGGGCCCTGGCCCCACCTCCCGAGTTTCTCATATATAAAGACAGT	1336
Qy	1888	TGCTTCTTCACTTGAATCAAGGGGACCTTGTGCTGTAACAAATCTTCTTCTTGAAGTGA	1947
Db	1337	TGCTTCTTCACTTGAATCAAGGGGACCTTGTGCTGTAACAAATCTTCTTGAAGTGA	1396
Qy	1948	AAAGTTAGCACTTCTCTTTGAAGGTGTGACGCTCAACAAAGGCTGTGAGAAACAAGGA	2007
Db	1397	AAAGTTAGCACTTCTCTTTGAAGGTGTGACGCTCAACAAAGGCTGTGAGAAACAAGGA	1456
Qy	2008	GGGGAGCATTAAGGGGCAAAACCTATCTCTGGGACAGATGATTTCTTAAGTCCAGATCATTAA	2067
Db	1457	GGGGAGCATTAAGGGGCAAAACCTATCTCTGGGACAGATGATTTCTTAAGTCCAGATCATTAA	1516
Qy	2068	CTAGCTCTTTCAGACTATCTTACATATAGTGGGGGAAAGAAACAGAGTCGAAAGAG	2127
Db	1517	CTAGCTCTTTCAGACTATCTTACATATAGTGGGGGAAAGAAACAGAGTCGAAAGAG	1576
Qy	2128	AAACGCTGATTTATACAGCAAGTAAGAAGTGAAGTGAAGACTGATGATTTCAACTTCTGG	2187
Db	1577	AAACGCTGATTTATACAGCAAGTAAGAAGTGAAGTGAAGACTGATGATTTCAACTTCTGG	1636
Qy	2188	TAGATGGCCACAACCCAGCCGCAAGGGCATCGAAACAACAGGGCCTGGGGCAACTATGCA	2247
Db	1637	TAGATGGCCACAACCCAGCCGCAAGGGCATCGAAACAACAGGGCCTGGGGCAACTATGCA	1696
Qy	2248	TGTGCAAAAGAGATTGGCTCAGAGTTGTGGGGTGAAGAGTCCAACTTGGGGGACCTCAAA	2307
Db	1697	TGTGCAAAAGAGATTGGCTCAGAGTTGTGGGGTGAAGAGTCCAACTTGGGGGACCTCAAA	1756
Qy	2308	TTATGTCTCTGGGTGATTTCAAGTAAACACCACTCATAGGCTTGTGTGTCATAGATTAGGCA	2367
Db	1757	TTATGTCTCTGGGTGATTTCAAGTAAACACCACTCATAGGCTTGTGTGTCATAGATTAGGCA	1816
Qy	2368	TGACAAGTGAATGAAGTTGAAGTGGGGGAAACAGAAATACACAGAGCTGTGTGACAGGC	2427
Db	1817	TGACAAGTGAATGAAGTTGAAGTGGGGGAAACAGAAATACACAGAGCTGTGTGACAGGC	1876
Qy	2428	AAGCTGAGAGAGAGAAAGAAATGAATTGCAACATGAGACCACTTTTGCAGAAACAGT	2487

Db	1877	AAGCTGAGAGAGAGAGAAAGAAATGAATGGACACATGGAGCACATTTGCAGAAACAGT	1336
Qy	2488	CCCTGGAGAGCTTCGCTGAGAGCTTCAGAGCTTTTGCTGGACAAGAGATCTGGGCTTACCCA	2547
Db	1937	CCCTGGAGAGCTTCGCTGAGAGCTTCAGAGAGCTTTTGCTGGACAAGAGATCTGGGCTTACCCA	1996
Qy	2548	ATTAGGCTCTCGGGTACTCTGCACACATCTAACAAGCAAAATGTACCTGGCAAGGAGTTGC	2697
Db	1997	ATTAGGCTCTCGGGTACTCTGCACACATCTAACAAGCAAAATGTACCTGGCAAGGAGTTGC	2056
Qy	2608	AGTACTGGTATATTTCTGCTCATTAACGTGTGAAGGCTTTGGGGTCCAAATTTGTCTGAC	2667
Db	2057	AGTACTGGTATATTTCTGCTCATTAACGTGTGAAGGCTTTGGGGTCCAAATTTGTCTGAC	2116
Qy	2668	AGCTGTTTAACTGGGAAATTGGGCGCTAGACTATAGTATGTCTCAGACAAGGCCCTTA	2727
Db	2117	AGCTGTTTAACTGGGAAATTGGGCGCTAGACTATAGTATGTCTCAGACAAGGCCCTTA	2176
Qy	2728	TTCTCTCACTGCTTTTCAACCCAGCTGAGGTTTGGAGGCTGGCTTTTCAAGCTCAAAA	2787
Db	2177	TTCTCTCACTGCTTTTCAACCCAGCTGAGGTTTGGAGGCTGGCTTTTCAAGCTCAAAA	2236
Qy	2788	AATAGCGCTGAGTTTCCAGAGAGGGCCCTTAATCTGAGTCTCCGTCTCAGGCTCATTT	2847
Db	2237	AATAGCGCTGAGTTTCCAGAGAGGGCCCTTAATCTGAGTCTCCGTCTCAGGCTCATTT	2296
Qy	2848	TCCTTTCTCTGTAATAATAGACACATGCCACCACCTTTCAGTGCACATGATATAGACTC	2907
Db	2287	TCCTTTCTCTGTAATAATAGACACATGCCACCACCTTTCAGTGCACATGATATAGACTC	2356
Qy	2908	AAACCCATCCTTGAATCTGTCTTTGGGAAAGGGGCTCTGAGAGTGAACCCACATGTGGCTC	2967
Db	2357	AAACCCATCCTTGAATCTGTCTTTGGGAAAGGGGCTCTGAGAGTGAACCCACATGTGGCTC	2416
Qy	2968	AATGGCCTCATGTATCTGAGAGTCAGGCCCTGCCAACCTGTTCAGGACATTTGCTCCGTAGGA	3027
Db	2417	AATGGCCTCATGTATCTGAGAGTCAGGCCCTGCCAACCTGTTCAGGACATTTGCTCCGTAGGA	2476
Qy	3028	CTTTGATATGGATAGAGTAGTAGTACTTAAACAAGCTCTGACTGTGCACACAAAGGCTTTGTACTGGG	3087
Db	2477	CTTTGATATGGATAGAGTAGTAGTACTTAAACAAGCTCTGACTGTGCACACAAAGGCTTTGTACTGGG	2536
Qy	3088	AGGCCAGGCTATAGAGTGGCTCCAGCTTTAAAAGGGCTGGAGCTGGGGGACAAGTGTCTCAG	3147
Db	2537	AGGCCAGGCTATAGAGTGGCTCCAGCTTTAAAAGGGCTGGAGCTGGGGGACAAGTGTCTCAG	2596
Qy	3148	ATTAGGGCTCTAATAGAAAGTTTGACTGAGCTGAGACAGAGGTTAAGGGGCAAGGACGA	3207
Db	2597	ATTAGGGCTCTAATAGAAAGTTTGACTGAGCTGAGACAGAGGTTAAGGGGCAAGGACGA	2656
Qy	3208	GGGTTGTGGGCTCTACTCCTTAGAGAGCACTTGAAGCTTTACTTTTCATTTCTCTAATGATGTC	3267
Db	2657	GGGTTGTGGGCTCTACTCCTTAGAGAGCACTTGAAGCTTTACTTTTCATTTCTCTAATGATGTC	2716
Qy	3268	TTGATATGGCTAACCTTCACAGGGGTTGGCTGACTAGTCTAAGGGGTTGAGACAAAGACAGAGT	3327
Db	2717	TTGATATGGCTAACCTTCACAGGGGTTGGCTGACTAGTCTAAGGGGTTGAGACAAAGACAGAGT	2776
Qy	3328	TTTCAAGTCTGGTCTTATCAAGTTCAAGTCAAGCACTTTGGGACCACTGCTGCATCAATGCC	3387
Db	2777	TTTCAAGTCTGGTCTTATCAAGTCTAAGTCAAGCACTTTGGGACCACTGCTGCATCAATGCC	2836
Qy	3388	AGGAGGCTTAAAGGATGCTAATAACAGTTATCCAAACAATGTGATACCAAGGTTAACTTTTC	3447
Db	2837	AGGAGGCTTAAAGGATGCTAATAACAGTTATCCAAACAATGTGATACCAAGGTTAACTTTTC	2896
Qy	3448	TCTGTTTTCAGAGGACAGGAGTACTAAGTCTCCCTTTCTCTTCTCTCCACGATGTC	3507
Db	2897	TCTGTTTTCAGAGGACAGGAGTACTAAGTCTCCCTTTCTCTTCTCTCCACGATGTC	2956
Qy	3508	TCTTGCAGAGGAATCTCTCTAGGTTGTCTTCAGAGGAATCTCCAGAAATAGGTTTGTTCAGTC	3567
Db	2957	TCTTGCAGAGGAATCTCTCTAGGTTGTCTTCAGAGGAATCTCCAGAAATAGGTTTGTTCAGTC	3016



Db 10181 AAAAAGATAGAGAACTAGGAGAGTCAGGTGTACAGAAAGCAAGTTCCAAAAA 10240  
Qy 241 GACATTTAAAGGAAAGAGTAGAGCAGTGCCTCTGAAAGGTACAGAT 300  
Db 10241 GACATTTAAAGGAAAGAGTAGAGCAGTGCCTCTGAAAGGTACAGAT 10300  
Qy 301 GAGACAGAGAAATGACATGAGATTTGCAATTGGAAGATCTAGCACTGATAG 360  
Db 10301 GAGACAGAGAAATGACATGAGATTTGCAATTGGAAGATCTAGCACTGATAG 10360  
Qy 361 AACAAATTCATGTTGAGGAGAAACAGAGTATTTGAAGAGATTGAGAAAAAGAC 420  
Db 10361 AACAAATTCATGTTGAGGAGAAACAGAGTATTTGAAGAGATTGAGAAAAAGAC 10420  
Qy 421 AAATGGAGCCTTGAATATTTCTTAAATAGTTGTTGAAAAAGAGAAACGGG 480  
Db 10421 AAATGGAGCCTTGAATATTTCTTAAATAGTTGTTGAAAAAGAGAAACGGG 10480  
Qy 481 GTGCTAGCCAGTACCTCCCTCACTCTCCACACCTCATAGGAGAGACTGGAGAAC 540  
Db 10481 GTGCTAGCCAGTACCTCCCTCACTCTCCACACCTCATAGGAGAGACTGGAGAAC 10540  
Qy 541 AGCCAGAGTGAACAATTCAGTGAAGTGTCTCTTTTAAATTTGAGCACTGTAT 600  
Db 10541 AGCCAGAGTGAACAATTCAGTGAAGTGTCTCTTTTAAATTTGAGCACTGTAT 10600  
Qy 601 TTCAATATCTAATCCGCACTCTCTGATGAGACCTGAAATCTTGAAGATGCCGC 660  
Db 10601 TTCAATATCTAATCCGCACTCTCTGATGAGACCTGAAATCTTGAAGATGCCGC 10660  
Qy 661 CAACCCATGATGTGGCCTTACCTGGAACCTTAGCCATGTTTCCACCTTGCTTC 720  
Db 10661 CAACCCATGATGTGGCCTTACCTGGAACCTTAGCCATGTTTCCACCTTGCTTC 10720  
Qy 721 TTTCAAGCACTGCTGATTCAGTTTCAGGAGGAGCAAGTGCCTCAATGTGACAA 780  
Db 10721 TTTCAAGCACTGCTGATTCAGTTTCAGGAGGAGCAAGTGCCTCAATGTGACAA 10780  
Qy 781 GTCTGTCTAATTTCTCTCTCTGACCTGGCTCTTCAATCTTGGCTCTGATTC 840  
Db 10781 GTCTGTCTAATTTCTCTCTCTGACCTGGCTCTTCAATCTTGGCTCTGATTC 10840  
Qy 841 CTCCTCTCATGACTCATTTCTGATTCATCACTAGCCCTCTCTGCTGGCTTCG 900  
Db 10841 CTCCTCTCATGACTCATTTCTGATTCATCACTAGCCCTCTCTGCTGGCTTCG 10900  
Qy 901 CAGCGGCTTGAAGCACTATGATTCAGAGGAGCCACTAAGTGCCTTGAACAC 960  
Db 10901 CAGCGGCTTGAAGCACTATGATTCAGAGGAGCCACTAAGTGCCTTGAACAC 10960  
Qy 961 AGCCGGGATGGCTGAGTACCTGAGCCATTTGCTCATAGTGGCCCGGCTCTTACGA 1020  
Db 10961 AGCCGGGATGGCTGAGTACCTGAGCCATTTGCTCATAGTGGCCCGGCTCTTACGA 11020  
Qy 1021 GTGAGTCTGCTTTCTGAGGAGAAACCTCATGCTGAGGAGGATTTGAGGAG 1080  
Db 11021 GTGAGTCTGCTTTCTGAGGAGAAACCTCATGCTGAGGAGGATTTGAGGAG 11080  
Qy 1081 GAGGCTTGAAGTGAAGCTGAGCAACAGATATCCACTGTATTTATCAATCTTTG 1140  
Db 11081 GAGGCTTGAAGTGAAGCTGAGCAACAGATATCCACTGTATTTATCAATCTTTG 11440  
Qy 1141 TCTCTCAGGAGCCTCAGATATATGCTTCTGACAAAGGAGGATTAACCTAGTAAGC 1200  
Db 11441 TCTCTCAGGAGCCTCAGATATATGCTTCTGACAAAGGAGGATTAACCTAGTAAGC 1200  
Qy 1201 GTTATCCGAAGGCTGAGAGAGAGTGGGAGCCCTCATAGGAGTATCTGAGACTG 1260  
Db 1201 GTTATCCGAAGGCTGAGAGAGAGTGGGAGCCCTCATAGGAGTATCTGAGACTG 1260  
Qy 1261 TGAATGCGGCTTATATCTGCTGGGTCTTCTGAGCTCAATATCATGAGTGAAGG 1320  
Db 11261 TGAATGCGGCTTATATCTGCTGGGTCTTCTGAGCTCAATATCATGAGTGAAGG 11320

Qy 1321 TTCTGGGTCTTGAAGGAGCTTGTCTTCTACCTGTCTGTGGCATAGTCCCAACG 1380  
Db 11321 TTCTGGGTCTTGAAGGAGCTTGTCTTCTACCTGTCTGTGGCATAGTCCCAACG 11380  
Qy 1381 GGCATGAGAGGCTTGAAGTCAAGATCCCAAGGAGTGAAGAGGCTTGAAGATCCC 1440  
Db 11381 GGCATGAGAGGCTTGAAGTCAAGATCCCAAGGAGTGAAGAGGCTTGAAGATCCC 11440  
Qy 1441 ATGACATGGAAGCCTATGTTGTTGTTGCTTCTTCCAGAGACGGCGCTGTGCT 1500  
Db 11441 ATGACATGGAAGCCTATGTTGTTGTTGCTTCTTCCAGAGACGGCGCTGTGCT 11500  
Qy 1501 GGAACCTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
Db 11501 GGAACCTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11560  
Qy 1561 AGACGAGGCTTGTATGAGAAAGTCCCTGGGCTTAACTGATTTCCGCAATGCTC 1620  
Db 11561 AGACGAGGCTTGTATGAGAAAGTCCCTGGGCTTAACTGATTTCCGCAATGCTC 11620  
Qy 1621 CGGCTTGAATCTATCATGTTGCCAATTTGATCTGTAAGTATGAGAGAACTGAA 1680  
Db 11621 CGGCTTGAATCTATCATGTTGCCAATTTGATCTGTAAGTATGAGAGAACTGAA 11680  
Qy 1681 TGAAGCAGAGGCTTGGGAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
Db 11681 TGAAGCAGAGGCTTGGGAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11740  
Qy 1741 AGGAGGCTTCTGACATGATCTGAGCTGAGCCCACTCTGATTTCTCTAAATGA 1800  
Db 11741 AGGAGGCTTCTGACATGATCTGAGCTGAGCCCACTCTGATTTCTCTAAATGA 11800  
Qy 1801 GATGCTTCTTGGCTTCTGATGAGGAGGCTTCTGATGAGTCTGAGGCTTCCCACT 1860  
Db 11801 GATGCTTCTTGGCTTCTGATGAGGAGGCTTCTGATGAGTCTGAGGCTTCCCACT 11860  
Qy 1861 CCCAGTTCTGATATTAAGACAGATGCTTCTTCACTGATCAAGGAGCCTTGGT 1920  
Db 11861 CCCAGTTCTGATATTAAGACAGATGCTTCTTCACTGATCAAGGAGCCTTGGT 11920  
Qy 1921 TGAACATCTTCTTCTTGAAGTGAAGATTAAGCACTTCTTGAAGGCTGAGC 1980  
Db 11921 TGAACATCTTCTTCTTGAAGTGAAGATTAAGCACTTCTTGAAGGCTGAGC 11980  
Qy 1981 TCAACAGGCTTGAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040  
Db 11981 TCAACAGGCTTGAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12040  
Qy 2041 AGATGAATCTTGAAGTCAAGATCAATAGCTTCTTGAAGTCAATAGGAGG 2100  
Db 12041 AGATGAATCTTGAAGTCAAGATCAATAGCTTCTTGAAGTCAATAGGAGG 12100  
Qy 2101 GGGAGAGAGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
Db 12101 GGGAGAGAGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 12160  
Qy 2161 AGCTAGAGCTTGAATCAATCTGCTGATGAGGAGCAACAGGAGGAGGAGGAGGAG 2220  
Db 12161 AGCTAGAGCTTGAATCAATCTGCTGATGAGGAGCAACAGGAGGAGGAGGAGGAG 12220  
Qy 2221 AACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2280  
Db 12221 AACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12280  
Qy 2281 AGAGGATCAATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340  
Db 12281 AGAGGATCAATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12340  
Qy 2341 ATGCTTGTGTTGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2400  
Db 12341 ATGCTTGTGTTGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 12400

Result	Query				Description	
No.	Score	Match	Length	DB	ID	
1	3737	100.0	13737	13	US-09-900-448-3	Sequence 3, Appl1
2	2279.2	61.0	3166	13	US-09-900-448-1	Sequence 1, Appl1
3	446	11.9	1631	15	US-10-123-237-19	Sequence 19, Appl1
4	446	11.9	1631	15	US-10-105-891-19	Sequence 19, Appl1
c	265.2	7.1	504	13	US-09-880-107-3166	Sequence 3166, Appl1
5	198.6	5.3	1516	13	US-10-316-253-39	Sequence 39, Appl1
6	198.6	5.3	1516	15	US-10-175-523-94	Sequence 84, Appl1
7	198.6	5.3	400	10	US-09-960-352-6580	Sequence 6580, Appl1
c	197.4	5.3	400	10	US-09-960-352-11316	Sequence 11316, Appl1
8	180.6	4.5	387	10	US-09-960-352-7040	Sequence 7040, Appl1
9	167	4.5	408	10	US-09-918-985-10754	Sequence 30754, Appl1
10	149.4	4.0	488	11	US-09-918-985-32181	Sequence 32181, Appl1
11	134.2	3.6	473	11	US-09-918-985-32181	Sequence 32820, Appl1
12	134.2	3.6	481	11	US-09-918-985-32820	Sequence 10323, Appl1
13	123.6	3.3	420	10	US-09-960-352-10323	Sequence 5368, Appl1
14	118.2	3.2	374	10	US-09-960-352-5368	
15						

16	117.4	3.1	383	10	US-09-960-352-6846	Sequence 6846, Ap
17	111.2	3.0	2288	10	US-09-960-352-8240	Sequence 8240, Ap
18	106.4	2.8	27469	13	US-10-017-161-1015	Sequence 1015, Ap
19	106	2.8	337	10	US-09-960-352-1236	Sequence 12936, A
20	105.6	2.8	661	13	US-10-027-633-136875	Sequence 136875, A
21	105.6	2.8	661	14	US-10-027-633-136875	Sequence 136875, A
22	105.4	2.8	617	13	US-10-027-632-102912	Sequence 102912, A
23	105.4	2.8	617	14	US-10-027-633-102912	Sequence 102912, A
24	105.4	2.8	337658	10	US-09-813-320-3	Sequence 3, Appl
25	103.2	2.8	203	10	US-09-960-352-9190	Sequence 9190, Ap
26	102.4	2.7	599	13	US-10-027-633-236392	Sequence 236392, A
27	102.4	2.7	599	13	US-10-027-632-236393	Sequence 236393, A
28	102.4	2.7	599	13	US-10-027-633-236394	Sequence 236394, A
29	102.4	2.7	599	14	US-10-027-632-236392	Sequence 236392, A
30	102.4	2.7	599	14	US-10-027-632-236393	Sequence 236393, A
31	102.4	2.7	5599	14	US-10-027-633-236394	Sequence 236394, A
32	102.4	2.7	2522	14	US-10-027-632-225590	Sequence 225590, A
33	102.4	2.7	2522	14	US-10-027-633-225590	Sequence 225590, A
34	102.2	2.7	343	10	US-09-960-352-12972	Sequence 12972, A
35	101	2.7	715	13	US-10-027-632-202762	Sequence 202762, A
36	101	2.7	715	13	US-10-027-633-202763	Sequence 202763, A
37	101	2.7	715	14	US-10-027-632-202762	Sequence 202762, A
38	101	2.7	715	14	US-10-027-633-202763	Sequence 202763, A
39	100.2	2.7	553	9	US-09-864-761-11839	Sequence 11839, A
40	99.2	2.7	650	13	US-10-027-632-240569	Sequence 240569, A
41	99.2	2.7	650	14	US-10-027-633-240569	Sequence 240569, A
42	99.2	2.7	988	13	US-10-027-633-122186	Sequence 122186, A
43	99.2	2.7	988	14	US-10-027-632-122186	Sequence 122186, A
44	99.2	2.7	1260	13	US-10-027-633-240568	Sequence 240568, A
45	99.2	2.7	1260	14	US-10-027-632-240568	Sequence 240568, A

## ALIGNMENTS

```

RESULT 1
US-09-900-448-3
; Sequence 3, Application US/09900448
; Publication No. US20030220486A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001272
; CURRENT APPLICATION NUMBER: US/09/900,448
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 13737
; TYPE: DNA
; ORGANISM: Human
; US-09-900-448-3

```

	Query Match	Best Local Similarity	Score 3737;	DB 13;	Length 13737;
	Matches 3737;	Conservative	0;	Mismatches	0; Gaps 0;
QY	1	CTCAAGAGAGAGGTCTAGCTCAGACTGACGAGAACAGATTGGAGTCAATCAGCATTTATGAT	60		
Db	10001	CTCAAGAGAGAGGTCTAGCTCAGACTGACGAGAACAGATTGGAGTCAATCAGCATTTATGAT	1006		
QY	61	GGCATTTTAAAGCCCTTGACTGACTAGTGGATTACCAAGAGAGTGAAGGTAGAGAGAGA	120		
Db	10061	GGCATTTTAAAGCCCTTGACTGACTAGTGGATTACCAAGAGAGTGAAGGTAGAGAGAGA	1012		
QY	121	GAAGAGGCCCAAGTGGGGGATTCACATATTTCATATCATAGTTGAGAAAGAGTATGTC	180		
Db	10121	GAAGAGGCCCAAGTGGGGGATTCACATATTTCATATCATAGTTGAGAAAGAGTATGTC	1018		
QY	181	AAAAAAGATTAAAGGAATATCTGGGAGAGTCAAGGTGTCACAGAGCCAGTTCCAAAAAAA	240		

EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 197  
LENGTH: 1282  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (675)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1195)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-205-258-197

Query Match 1.9%; Score 69.2; DB 4; Length 1282;  
Best Local Similarity 73.0%; Pred. No. 1.6e-11;  
Matches 89; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 3554 GCTTGTTCAGTCTGCTGCTGCTATGAGAGAAATCTTAGAGTGGTAACTATCA 3613  
DB 505 GGTATGGTATCTTCTTGTGCTGCTATACAGAAATACAGAGACTGGGTAATTGTAA 446  
QY 3614 GCATAGCAATTTATTTGTCACATCTCGAGGCTGGAATAATCCAAAGATCAAGGCTCAG 3673  
DB 445 GGAACAGAAATGTAATTTCTACACAGTTCTGAGAGATGGAGAGTCTTAAGATCAAGAGCTGG 386  
QY 3674 CA 3675  
DB 385 CA 384

Search completed: December 16, 2003, 05:46:27  
Job time: 161.331 secs

;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089908  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089948  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089952  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090246  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090252  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090254  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090355  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090431  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090435  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090444  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090472  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090535  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090540  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090542  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090676  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090678  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090690  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090694  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090695  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090696  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090862  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091478  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091544  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091626  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091633  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07

;; PRIOR APPLICATION NUMBER: 60/092182  
;; PRIOR FILING DATE: 1998-07-09

Query Match 1.9%; Score 69.8; DB 4; Length 762;

Best Local Similarity 71.3%; Pred. No. 7e-12; Matches 92; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 3609 TATCAGCAATAGAGATTTATTTATTTGTTCAATTTCTGAGGCTGGAATTCAGATCAAGGC 3668

Db 756 TATTAACAGCAATATTTATTTATTTCTTACAGTTCTAGAGGCTGGAGAGTCCAGATCAAGGC 697

Qy 3669 TCCAGCAGGTTCAAGTCTGCTGAGTGTGTTCTTCTGAGATGACACCTTTTGTCT 3728

Db 696 ACCAGCAGATTTCAAGCATCTGTGTGAGACCAAGTTTCTTCTCTGACAGTGCCCTTAGCT 637

Qy 3729 GTGTTCTCA 3737

Db 636 GTGATCACA 628

# RESULT 15

US-09-205-258-197/c  
; Sequence 197, Application US/09205258

; Patent No. 6525174

; GENERAL INFORMATION:

; APPLICANT: Young et al.

; TITLE OF INVENTION: 207 Human Secreted Proteins

; FILE REFERENCE: P2007P1

; CURRENT APPLICATION NUMBER: US/09/205,258

; EARLIER FILING DATE: 1998-12-04

; EARLIER APPLICATION NUMBER: PCT/US98/11422

; EARLIER FILING DATE: 1998-06-04

; EARLIER APPLICATION NUMBER: 60/048,885

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,375

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,881

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,880

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,876

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,895

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,884

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,894

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,971

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,964

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,893

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,900

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,901

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,892

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,915

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,019

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,970

Db 1184 YY 1243  
Qy 744 TTTCAGCAGGAGGACAGTGCACATTCGACCAAGTGTCTATTTCTCTC 803  
Db 1244 YY 1303  
Qy 804 ACCTGCGCTCTTCATCTTGCCTGTGATGATCTCTCTCTCATGACTGATTCG 863  
Db 1304 YY 1363  
Qy 864 CATTCATCTAGCTCTTCTCTGCGCTGTGATGATCTCTCTCTCATGACTGATG 923  
Db 1364 YY 1423  
Qy 924 GTATTCACAGGAGGACCACTAGTGGCTGTGACACAGCGGGATGCTG 975  
Db 1424 YYYYYYYYYYGGACCAATTCCTATCTTAACTTACTGATGATG 1475

RESULT 14  
US-09-996-243-344/C  
Sequence 344, Application US/09996243  
Patent No. 6478825  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertlisen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: KJavian, Ivar J.  
APPLICANT: Nadler, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2730P1C13  
CURRENT APPLICATION NUMBER: US/09/996,243  
PRIOR APPLICATION NUMBER: 2001-11-14  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/075945  
PRIOR FILING DATE: 1998-02-25  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28

PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088021  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088030  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088326  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088202  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088212  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088734  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088742  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089440  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801



Query Match 2.1%; Score 78.4; DB 3; Length 8353;

Best Local Similarity 66.8%; Pred. No. 6.5e-14; Indels 2; Gaps 1;

Matches 127; Conservative 0; Mismatches 61; Indels 2; Gaps 1;

QY 3547 CAGAAATGGTTTGTTCAGTCAGTTAGGCTGCTATAGAAATATCTTAGAGTGGTAA 3606

DB 8244 CAGTATTTGTATTTGTCTTAGCTGGAGTATTAACAAATATACAGCAGCTGGGTG 8185

QY 3607 TCATACGCAATAGCAATTTATTTGTTCACATTTCTGAGGCTGGAATTCAGATTAAG 3666

DB 8184 CTTATTAACATAGAAATGTCTTCTTAGAGTCTGAGGAGTGGAAATCAAGATCAAT 8125

QY 3667 GCTCCAGAGTTCAGTGTGCTGAGTGTCTGT--CTGCTGGAAGTGGACCTTTT 3724

DB 8124 GCACCAACAGCTTTGTGTCTGTGAGGCGCAGTTTCTCTGATATGATGCTTTGT 8065

QY 3725 TGCTGTGTTT 3734

DB 8064 CACTGGGCTC 8055

RESULT 12

US-09-009-913-1/c

Sequence 1, Application US/0909913

Patent No. 6087485

GENERAL INFORMATION:

APPLICANT: Axy's Pharmaceuticals, Inc.

TITLE OF INVENTION: Asthma Related Genes

NUMBER OF SEQUENCES: 339

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bozicevic & Reed, LLP

STREET: 285 Hamilton Ave, Suite 200

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/009,913

FILING DATE: 21-Jan-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J

REGISTRATION NUMBER: 36,677

REFERENCE/DOCKET NUMBER: SEQ-4P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-327-3231

TELEFAX: 650-327-3231

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 72928 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-009-913-1

Query Match 2.1%; Score 77.6; DB 3; Length 72928;

Best Local Similarity 72.2%; Pred. No. 5.6e-13; Indels 6; Gaps 2;

Matches 130; Conservative 0; Mismatches 44; Indels 6; Gaps 2;

QY 3560 TTTCAGTCAAGTTAGGCTGCTATAGCAATATCTTAGAGTGGTATCTATCAGCAATA 3619

DB 3623 TCTTATTCATTAACGCTGCACAAAAGATGCTATGACTAGS----CTACAGCAAGC 36208

QY 3620 GGAATTTATGTTTACATTTCTGAGCTGGAATATCAAGATCAGGCTCGCAGAGTT 3679

DB 36207 GAAATTTATTTATTTCAAAATTTCTAGAGCTGGAGTCCAGAGCAAGCAGCAGATT 36148

QY 3680 CAGTGTCTGCTAGTGS--CTGTCTGCTTGAAGATGCGACCTTTTGTGTCTCA 3737

DB 36147 CAGCGTCTGTGAGGCGCCACCTTCTTAATGATGATGATCTTCTCACCGTACCTCA 36088

RESULT 13

US-08-232-463-14

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHREIFLINGER, F.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-9300

TELEFAX: (703) 683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PTZ9PT-F18

US-08-232-463-14

Query Match 1.9%; Score 70; DB 1; Length 7218;

Best Local Similarity 5.3%; Pred. No. 3e-11;

Matches 22; Conservative 235; Mismatches 155; Indels 0; Gaps 0;

QY 564 GAAGTGTCTTCCTTTTAACTTGAACACTGATTTCAATTAATCAACCGCATCTC 623

DB 1064 GATT 1123

QY 624 TGACATGACACCTGAATCTTATAGGAGTGCAGCCCAACCCAGATGTTGCGCTTAC 683

DB 1124 YY 1183

QY 684 CTGGAACCTTAGCACTGTTTCCACACTGCTTCTTCAAGCACTGCTGATTCAG 743



OTHER INFORMATION: 12-500-220 : polymorphic base A or G  
NAME/KEY: misc binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-500-220.mis1, potential  
NAME/KEY: misc binding  
LOCATION: 502..521  
OTHER INFORMATION: 12-500-220.mis2, potential complement  
NAME/KEY: primer bind  
LOCATION: 283..303  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer bind  
LOCATION: 711..731  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc binding  
LOCATION: 489..513  
OTHER INFORMATION: 12-500-220 potential probe  
US-09-671-317-212

Query Match 2.1%; Score 79; DB 4; Length 1001;  
Best Local Similarity 75.5%; Pred. No. 9.1e-15;  
Matches 111; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 3558 TGTTCAGTCAGTTAGGCTGCTATAGA-GAATATCTTAGTGGTAACTATCAGCA 3616  
DB 837 TGTCTGTGTTGTTGGGCTGCTGTAACAAATACCTTACAGTGGCTGCTATAACA 896  
QY 3617 ATAGAAATTATTTGTCACAAATTCGAGGCTGGAATCAAGATCAAGGCTCCAGCAG 3676  
DB 897 ACAGAAATGATTTGCTCACAAGTTCTGAGGCTGGGAAGTTCAAGATCAAGTACCGGCGAG 956  
QY 3677 GTTCAGTGTCTGCTGAGTGTCTTCT 3703  
DB 957 TTTCGGTGTGTGTGAGAGCTTTTGTCT 983

RESULT 10  
US-09-671-317-458  
Sequence 458, Application US/09671317  
Patent No. 6528260  
GENERAL INFORMATION:  
APPLICANT: Blumentfeld, Marra  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM  
FILE REFERENCE: 62.US3.CIP  
CURRENT APPLICATION NUMBER: US/09/671.317  
CURRENT FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US 09/536,178  
PRIOR FILING DATE: 2000-03-23  
PRIOR APPLICATION NUMBER: PCT/IB00/00403  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: US 60/126,269  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/131,961  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 977  
SOFTWARE: Patent.pm  
SEQ ID NO 458  
LENGTH: 1001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 12-500-217 : insertion CATATA  
NAME/KEY: misc binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-500-217.mis1, potential  
NAME/KEY: primer bind  
LOCATION: 286..306  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer bind

LOCATION: 714..734  
OTHER INFORMATION: downstream amplification primer, complement  
US-09-671-317-458

Query Match 2.1%; Score 79; DB 4; Length 1001;  
Best Local Similarity 75.5%; Pred. No. 9.1e-15;  
Matches 111; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 3558 TGTTCAGTCAGTTAGGCTGCTATAGA-GAATATCTTAGTGGTAACTATCAGCA 3616  
DB 840 TGTCTGTGTTGTTGGGCTGCTGTAACAAATACCTTACAGTGGCTGCTATAACA 899  
QY 3617 ATAGAAATTATTTGTCACAAATTCGAGGCTGGAATCAAGATCAAGGCTCCAGCAG 3676  
DB 900 ACAGAAATGATTTGCTCACAAGTTCTGAGGCTGGGAAGTTCAAGATCAAGTACCGGCGAG 959  
QY 3677 GTTCAGTGTCTGCTGAGTGTCTTCT 3703  
DB 960 TTTCGGTGTGTGTGAGAGCTTTTGTCT 986

RESULT 11  
US-08-611-587-1/c  
Sequence 1, Application US/08611587  
Patent No. 6150091  
GENERAL INFORMATION:  
APPLICANT: PANDOLFO, MASSIMO  
APPLICANT: MONTERMINI, LAURA  
APPLICANT: MOLTO, MARIA D.  
APPLICANT: Koenig, Michael  
APPLICANT: Campuzano, Victoria  
APPLICANT: Cossee, Mireille  
TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Fulbright & Jaworski L.L.P. Patent Dept.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.  
ZIP: 77010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,587  
FILING DATE: 03-MAR-1996  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Brashers-Macabee, Sarah J.  
REGISTRATION NUMBER: 38,087  
REFERENCE/DOCKET NUMBER: D-5901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5620  
TELEFAX: 713-651-5246  
TELEX: 76-2829  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8353 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 9q13  
UNITS: bp  
US-08-611-587-1

SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 174493  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)....(174493)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-804-471A-3

Query Match 2.2%; Score 82.6; DB 4; Length 174493;  
Best Local Similarity 72.2%; Pred. No. 2.6e-14;  
Matches 135; Conservative 0; Mismatches 49; Indels 3; Gaps 2;

QY 3551 AATGGTTGTTGACGAGTTAGGCTGCTATAGAGAAATCTTAAGTGGGTAATCTA 3610  
DB 14097 AAAGTGCATCTGATCTGTTGGCTGCGATATAG-ATACCTTAGACTTGGCAATTTA 14155  
QY 3611 TCAGCAATAGGAATTTATTTGTTCAATCTGAGGCTGGAATAATCAAGATCAAGGCTC 3670  
DB 14156 TAAACATATAGAAATTTCTTCTGACAGTTGTGAAGACTGGGAATCCAAAGTCAAGGCGC 14215  
QY 3671 CAGCAGGTTGAGTCTGCTGAGTGTGTTGCTTGAAGATGGACCTTTTGTCTGT 3730  
DB 14216 CAGCAGATCTGATCTGTGATGAGTCT--CCCTGCTTCAAAAATGGCGCTTCTGTCTGC 14273  
QY 3731 GTTCTCA 3737  
DB 14274 ATCTTCA 14280

## RESULT 7

US-09-873-404-3  
; Sequence 3, Application US/09873404  
; Patent No. 6500656  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL001212-CIP  
; CURRENT APPLICATION NUMBER: US/09/873,404  
; CURRENT FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 63588  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)....(63588)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-873-404-3

Query Match 2.2%; Score 80.8; DB 4; Length 63588;  
Best Local Similarity 71.6%; Pred. No. 4.7e-14;  
Matches 106; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 3550 AATGGTTGTTGACGAGTTAGGCTGCTATAGAGAAATCTTAAGTGGGTAATCT 3609  
DB 15658 AATTCATCTGTTTATGCTATTCAGAGCATATACAAATACATTAAGTGGCTT 15717  
QY 3610 ATCAGCAATAGGAATTTATTTGTTCAATCTGAGGCTGGAATAATCAAGATCAAGGCT 3669  
DB 15718 ATAAACAGCAGAAATTTATTTCTGAGGCTTCTGAGGCTGGAAAGTCAAGATCAAGGCTG 15777  
QY 3670 CAGCAGGTTGAGTCTGCTGAGTGTCT 3697  
DB 15778 CAGGCAATTCAGTGTCTGAGGCTT 15805

## RESULT 8

US-09-146-053-4  
; Sequence 4, Application US/09146053A  
; Patent No. 6399349  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, James W.  
; APPLICANT: Sprinkle, Terry Joe Curtis  
; APPLICANT: Venema, Richard C.  
; TITLE OF INVENTION: Human Aminopeptidase P Gene  
; FILE REFERENCE: MCG103  
; CURRENT APPLICATION NUMBER: US/09/146,053A  
; CURRENT FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/057,854  
; EARLIER FILING DATE: 1997-09-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 50000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-146-053-4

Query Match 2.2%; Score 80.4; DB 4; Length 50000;  
Best Local Similarity 68.7%; Pred. No. 5.4e-14;  
Matches 125; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 3556 TTGTTTCACTGCTAGTCTGCTGCTATAGAGAAATCTTAAGTGGGTAATCTATCAGC 3615  
DB 37896 TTATCTTATGATCTGAGGCTGCTATTAACAAAGGCTGCAAGCTGTTGCTTATTAAC 37955  
QY 3616 AATAGAAATTTATTTGTTCAATCTGAGGCTGGAATAATCAAGATCAAGGCTCAGCA 3675  
DB 37956 AGTAGACATTTATTTTTCATGCTGAGGCTGGAATAATCAAGATCAAGGCTCAGCA 38015  
QY 3676 GGTTCAGTCTGCTGAGTGTGTTGCTTGAAGATGGACCTTTTGTCTGTCTGT 3735  
DB 38016 GATTGTGTCCAAAGGCGCCAGT-TCCTTGTATGATGGACCTCTAGCTGTATCT 38074  
QY 3736 CA 3737  
DB 38075 CA 38076

## RESULT 9

US-09-671-317-212  
; Sequence 212, Application US/09671317  
; Patent No. 6528260  
; GENERAL INFORMATION:  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Chumakov, Ilya  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Cohen, Amick  
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM  
; FILE REFERENCE: 62,US3,CIP  
; CURRENT APPLICATION NUMBER: US/09/671,317  
; CURRENT FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US 09/536,178  
; PRIOR FILING DATE: 2000-03-23  
; PRIOR APPLICATION NUMBER: PCT/IB00/00403  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: US 60/126,269  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/131,961  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 977  
; SOFTWARE: Patent.pm  
; SEQ ID NO 212  
; LENGTH: 1001  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: allele  
; LOCATION: 501

1	OTHER INFORMATION:	unknown
2	NAME/KEY: unsure	
3	LOCATION: 29981	
4	OTHER INFORMATION:	unknown
5	NAME/KEY: unsure	
6	LOCATION: 30136	
7	OTHER INFORMATION:	unknown
8	NAME/KEY: unsure	
9	LOCATION: 30140	
10	OTHER INFORMATION:	unknown
11	NAME/KEY: unsure	
12	LOCATION: 31205	
13	OTHER INFORMATION:	unknown
14	NAME/KEY: unsure	
15	LOCATION: 31206	
16	OTHER INFORMATION:	unknown
17	NAME/KEY: unsure	
18	LOCATION: 31592	
19	OTHER INFORMATION:	unknown
20	NAME/KEY: unsure	
21	LOCATION: 33095	
22	OTHER INFORMATION:	unknown
23	NAME/KEY: unsure	
24	LOCATION: 33160	
25	OTHER INFORMATION:	unknown
26	NAME/KEY: unsure	
27	LOCATION: 34056	
28	OTHER INFORMATION:	unknown
29	NAME/KEY: unsure	
30	LOCATION: 34072	
31	OTHER INFORMATION:	unknown
32	NAME/KEY: unsure	
33	LOCATION: 36816	
34	OTHER INFORMATION:	unknown
35	NAME/KEY: unsure	
36	LOCATION: 39020	
37	OTHER INFORMATION:	unknown
38	NAME/KEY: unsure	
39	LOCATION: 42154	
40	OTHER INFORMATION:	unknown
41	NAME/KEY: unsure	
42	LOCATION: 42459	
43	OTHER INFORMATION:	unknown
44	NAME/KEY: unsure	
45	LOCATION: 46808	
46	OTHER INFORMATION:	unknown
47	NAME/KEY: unsure	
48	LOCATION: 46823	
49	OTHER INFORMATION:	unknown
50	NAME/KEY: unsure	
51	LOCATION: 46826	
52	OTHER INFORMATION:	unknown
53	NAME/KEY: unsure	
54	LOCATION: 47221	
55	OTHER INFORMATION:	unknown
56	NAME/KEY: unsure	
57	LOCATION: 52786	
58	OTHER INFORMATION:	unknown
59	NAME/KEY: unsure	
60	LOCATION: 52787	
61	OTHER INFORMATION:	unknown
62	NAME/KEY: unsure	
63	LOCATION: 53384	
64	OTHER INFORMATION:	unknown
65	NAME/KEY: unsure	
66	LOCATION: 54684	
67	OTHER INFORMATION:	unknown
68	NAME/KEY: unsure	
69	LOCATION: 59215	
70	OTHER INFORMATION:	unknown
71	NAME/KEY: unsure	
72	LOCATION: 59225	
73	OTHER INFORMATION:	unknown

1	NAME/KEY: unsure	LOCATION: 59242	
2	OTHER INFORMATION:	unknown	
3	NAME/KEY: unsure	LOCATION: 63280	
4	OTHER INFORMATION:	unknown	
5	NAME/KEY: unsure	LOCATION: 66614	
6	OTHER INFORMATION:	unknown	
7	NAME/KEY: unsure	LOCATION: 68650	
8	OTHER INFORMATION:	unknown	
9	NAME/KEY: unsure	LOCATION: 68697	
10	OTHER INFORMATION:	unknown	
11	NAME/KEY: unsure	LOCATION: 68718	
12	OTHER INFORMATION:	unknown	
13	NAME/KEY: unsure	LOCATION: 68733	
14	OTHER INFORMATION:	unknown	
15	NAME/KEY: unsure	LOCATION: 68729	
16	OTHER INFORMATION:	unknown	
17	NAME/KEY: unsure	LOCATION: 69785	
18	OTHER INFORMATION:	unknown	
19	NAME/KEY: unsure	LOCATION: 79134	
20	OTHER INFORMATION:	unknown	
21	NAME/KEY: unsure	LOCATION: 79188	
22	OTHER INFORMATION:	unknown	
23	NAME/KEY: unsure	LOCATION: 86336	
24	OTHER INFORMATION:	unknown	
25	OTHER INFORMATION:	unknown	
26	OTHER INFORMATION:	unknown	

Query Match	2.3%;	Score 86;	DB 4;	Length 87543;
Best Local Similarity	73.3%;	Pred. No. 1.2e-15;		
Matches 151;	Conservative	0;	Mismatches 50;	Indels 5;
				Gaps 3;

QY	3534	TTCCAGGGAACTCCAGAAATGGTTTGTTCAGTCAGTAAAGCTGTG---ATPAAGAAAT	3590
Db	68969	TACTAGTAATGCCAGAACTGAATTTGTCTTAAGTCTGTCTTTGCTGTTTCATACAAAT	68910
QY	3591	ATCTTAGAGTGGTAAATCTATCAGCAATAGAAATTATTTGCATCAATTCGAGAGCTGG	3650
Db	68909	GCCACAGACTGGGCATTTTACAAACAAAGTAGTTATTTTCATAGTTCTGGAAGCTGG	68850
QY	3651	AAAATCCAGATCAAGGCTTCAGACAGTTCAAGTGTCTGCTAGAGCTTTGT-TCGTGTTG	3709
Db	68849	-AGGTCACAAATCAAGGCGCCAGCTAGTTCAAGTGTCTGTGAGGCGCTGCTCTCTGTTCC	68791
QY	3710	AAGATGACACTTTTGTGCTGTGTTCT	3735
Db	68790	AAATGAGGCTGTGCTGTGTCTCT	68765

```

RESULT 6
US-09-804-471A-3
; Sequence 3, Application US/09804471A
; Patent No. 6479269
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01164
; CURRENT APPLICATION NUMBER: US/09/804,471A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 4

```

CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620lenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 682-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-79

Query Match 2.3%; Score 86; DB 3; Length 87350;  
Best Local Similarity 73.3%; Pred. No. 1.2e-15;  
Matches 151; Conservative 0; Mismatches 50; Indels 5; Gaps 3;

QY 3534 TCCAGGGAATCCAGAAATGTTTGTTCAGTCAAGTTAGGCTGCT--ATAAGAAAT 3590  
DB 68776 TACTAGTATGTCAGACGATGATTTGCTTAGTCTGTTTGTCTTTCATACAAAT 68717  
QY 3591 ATCTTAGAGTGGTAATCTATCAGCAATAGAAATTTATGTTCAATTCTGGAGGCTGG 3650  
DB 68716 GCCACAGACTGGGCAATTTACAAACACAGTAGTTATTTCTCATAGTTCTGGAGGCTGG 68657  
QY 3651 AATATCCAGATCAAGGCTCCAGAGGTTCAAGTGTCTGCTGAGTCTTGT-TCTGCTTGC 3709  
DB 68656 -AGTCCAGATCAAGGCGCCAGTAGGTTCAAGTGTCTGAGAGGCTGCTCTGCTTCC 68598  
QY 3710 AAGATGGACCTTTTGTCTGTCTTCT 3735  
DB 68597 AATATGAGGCTTGTCTGTGTCTTCT 68572

## RESULT 4

US-09-618-166-79/C  
Sequence 79, Application US/09618166  
Patent No. 6583112  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
Oshima, Junko  
Mulligan, John T.  
Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/618,166  
FILING DATE: 17-Jul-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052.419C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 79:  
US-09-618-166-79

Query Match 2.3%; Score 86; DB 4; Length 87350;  
Best Local Similarity 73.3%; Pred. No. 1.2e-15;  
Matches 151; Conservative 0; Mismatches 50; Indels 5; Gaps 3;

QY 3534 TCCAGGGAATCCAGAAATGTTTGTTCAGTCAAGTTAGGCTGCT--ATAAGAAAT 3590  
DB 68776 TACTAGTATGTCAGACGATGATTTGCTTAGTCTGTTTGTCTTTCATACAAAT 68717  
QY 3591 ATCTTAGAGTGGTAATCTATCAGCAATAGAAATTTATGTTCAATTCTGGAGGCTGG 3650  
DB 68716 GCCACAGACTGGGCAATTTACAAACACAGTAGTTATTTCTCATAGTTCTGGAGGCTGG 68657  
QY 3651 AATATCCAGATCAAGGCTCCAGAGGTTCAAGTGTCTGCTGAGTCTTGT-TCTGCTTGC 3709  
DB 68656 -AGTCCAGATCAAGGCGCCAGTAGGTTCAAGTGTCTGAGAGGCTGCTCTGCTTCC 68598  
QY 3710 AAGATGGACCTTTTGTCTGTCTTCT 3735  
DB 68597 AATATGAGGCTTGTCTGTGTCTTCT 68572

## RESULT 5

US-09-791-211-3/C  
Sequence 3, Application US/09791211  
Patent No. 6448080  
GENERAL INFORMATION:  
APPLICANT: Donna T. Ward  
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION  
FILE REFERENCE: RTS-0205  
CURRENT APPLICATION NUMBER: US/09/791,211  
CURRENT FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 90  
SEQ ID NO 3  
LENGTH: 87543  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 7421  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 7427  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 11609  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 12605  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 12742  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29370  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29422  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29979  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 29980

Best Local Similarity 7.5%; Pred. No. 2.6e-19;  
Matches 32; Conservative 249; Mismatches 144; Indels 0; Gaps 0;

Qy	93	CCAAGAAAGTCAMAGGTAGAGAGAGAAGAAAGAGCGCCCAAGTAGGGATTCATAATTT	152
Db	1471	CTATGCAAGTAGTTAAAGAGATGAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRR	1412
Qy	153	AGATATCAGGTTGAAGAAAAGATGACTCAAAAAAGATAAGAGAAATCTGGAGAGTCAG	212
Db	1411	RRR	1352
Qy	213	GTCACAGAAAGCCAGTTCCAAAAAAGACATTAAGAGAGAAAGTAGTAGCAGT	272
Db	1351	RRR	1292
Qy	273	CGAGTGTCTCTGAGAGTAGAGGTCCAGATGAGAACAGAAATTGACATGAGATTTCCGA	332
Db	1291	RRR	1232
Qy	333	ATTGAGAAATCTAGCAACCTCGATTAAGAACATTTCAATGTTAGGAAAAAGAAAGTG	392
Db	1231	RRR	1172
Qy	393	TAAATGAAAGAGATTGAGAAAAAAGAACAAATGAGCGCTAGATATCTTAATAAGTT	452
Db	1171	RRR	1112
Qy	453	GTTGTAAAAAGAGAGAAAGAAACGGGGTGTAGCCAGCTACTCTCACTTTCCAC	512
Db	1111	RRR	1052
Qy	513	CACCT 517	
Db	1051	GACCT 1047	

```

RESULT 2
US-09-539-333D-177
; Sequence 177, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguetelrec, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Esseloux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.0472AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 177
;
; LENGTH: 3001

```

```

TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 1501
OTHER INFORMATION: 99-15668-139 : polymorphic base C or T
FEATURE:
NAME/KEY: misc_binding
LOCATION: 1482..1500
OTHER INFORMATION: 99-15668-139.mtst1
FEATURE:
NAME/KEY: misc_binding
LOCATION: 1502..1521
OTHER INFORMATION: 99-15668-139.mts2, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1363..1380
OTHER INFORMATION: upstream amplification primer
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1801..1821
OTHER INFORMATION: downstream amplification primer, complement
FEATURE:
NAME/KEY: misc_binding
LOCATION: 1488..1513
OTHER INFORMATION: 99-15668-139 probe
US-09-539-333D-177

```

RESULT 3  
 US-08-781-891-79/c  
 Sequence 79, Application US/08781891  
 Patent No. 6090620  
 GENERAL INFORMATION:  
 APPLICANT: Fu, Ying-Hui  
 APPLICANT: Yu, Chang-Bn  
 APPLICANT: Oshima, Junko  
 APPLICANT: Mulligan, John T.  
 APPLICANT: Schellenberg, Gerald D.  
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
 NUMBER OF SEQUENCES: 209  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED AND BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/781,891  
 FILING DATE: 27-DEC-1996

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 11:21:38 ; Search time 158.331 Seconds  
(without adjustments)

10417.750 million cell updates/sec

Title: US-09-900-448-3\_COPY\_10001\_13737  
Page: 3737

Sequence: 1 ctcaaggaaggtcagtca.....acctttgtgctgtctca 3737

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database : Issued Patents NA: \*

```

1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfile1.seq.*

```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	95	2.5	7218	1	US-08-232-63-14	Sequence 14, Appl
C 2	88	2.4	3001	4	US-09-539-3330-177	Sequence 177, Appl
C 3	85	2.3	87350	3	US-08-781-891-79	Sequence 79, Appl
C 4	86	2.3	87350	4	US-09-618-166-79	Sequence 79, Appl
C 5	86	2.3	87543	4	US-09-791-211-3	Sequence 3, Appl1
C 6	82.6	2.2	174493	4	US-09-804-471-13	Sequence 3, Appl1
C 7	80.8	2.2	63588	4	US-09-873-404-3	Sequence 3, Appl1
C 8	80.4	2.2	50000	4	US-09-146-053-4	Sequence 4, Appl1
C 9	79	2.1	1001	4	US-09-671-317-212	Sequence 212, Appl
C 10	79	2.1	1001	4	US-09-671-317-458	Sequence 458, Appl
C 11	78.4	2.1	8353	3	US-08-611-587-1	Sequence 1, Appl1
C 12	77.6	2.1	72928	3	US-09-009-913-1	Sequence 1, Appl1
C 13	70	1.9	7218	4	US-08-232-63-14	Sequence 14, Appl
C 14	69.8	1.9	762	4	US-09-996-243-344	Sequence 344, Appl
C 15	69.2	1.9	1282	4	US-09-205-258-197	Sequence 197, Appl
C 16	69	1.8	50000	4	US-09-146-053-3	Sequence 3, Appl1
C 17	68.4	1.8	137496	4	US-09-877-177-10	Sequence 10, Appl1
C 18	68.2	1.8	309	4	US-09-222-575-145	Sequence 145, Appl
C 19	68.2	1.8	309	4	US-09-389-681-145	Sequence 145, Appl
C 20	68.2	1.8	309	4	US-09-620-405-145	Sequence 145, Appl
C 21	68.2	1.8	309	4	US-09-339-338-145	Sequence 145, Appl
C 22	68.2	1.8	309	4	US-09-433-826-145	Sequence 145, Appl
C 23	68.2	1.8	309	4	US-09-604-287-145	Sequence 145, Appl
C 24	67.8	1.8	50000	4	US-09-146-053-4	Sequence 4, Appl1
C 25	66.8	1.8	41684	4	US-09-536-059-1	Sequence 1, Appl1
C 26	66.6	1.8	1001	4	US-09-641-638-459	Sequence 459, Appl
C 27	63.2	1.7	246240	2	US-08-724-394-20	Sequence 20, Appl

C 28	63.2	1.7	246240	2	US-08-724-394-21	Sequence 21, Appl
C 29	63.2	1.7	246240	2	US-08-724-394-22	Sequence 22, Appl
C 30	63	1.7	3001	4	US-09-539-3330-1	Sequence 192, Appl
C 31	62.8	1.7	44453	4	US-09-146-053-5	Sequence 5, Appl
C 32	62	1.7	168575	4	US-09-426-3290-1	Sequence 1, Appl
C 33	62	1.7	339608	4	US-09-439-3330-1	Sequence 1, Appl
C 34	62	1.7	339608	4	US-09-679-409-1	Sequence 1, Appl
C 35	61.8	1.7	132331	3	US-09-128-155-16	Sequence 16, Appl
C 36	61.8	1.7	176373	3	US-09-128-155-17	Sequence 17, Appl
C 37	61.4	1.6	8355	3	US-08-406-0304-23	Sequence 23, Appl
C 38	60.4	1.6	3001	4	US-09-539-3330-13	Sequence 138, Appl
C 39	59.4	1.6	202001	4	US-09-734-694-3	Sequence 3, Appl
C 40	59	1.6	80246	3	US-09-078-2974-4	Sequence 4, Appl
C 41	59	1.6	80595	3	US-09-078-294-3	Sequence 3, Appl
C 42	57	1.5	49312	4	US-09-671-317-485	Sequence 485, Appl
C 43	55	1.5	116592	3	US-09-818-512-3	Sequence 3, Appl
C 44	53.6	1.4	49136	3	US-09-822-869-1	Sequence 1, Appl
C 45	52	1.4	43069	4	US-09-292-5424-1	Sequence 1, Appl

## ALIGNMENTS

[illegible]

THIS PAGE BLANK (USPTO)

CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
 CC  
 SQ Sequence 228 BP; 53 A; 58 C; 60 G; 57 T; 0 other;  
 Query Match 3.0%; Score 111.2; DB 25; Length 228;  
 Best Local Similarity 79.9%; Pred. No. 1.3e-23;  
 Matches 131; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
 QY 1158 GATATGCTTCTCTGACAAAGGAGGCTATACCTAGTAGCGGTTATCCGAAGCGGCTG 1217  
 DB 1 GTATATATCTTCTCTGACAAAGGAGGCTATACCTAGTAGCGGTTATCCGAAGCGGCTG 60  
 QY 1218 GAGAAGGAACTCGGACCCCTCATGGAATTATCTGGAAGCTGATGCGGCTTTATC 1277  
 DB 61 GAGAAGGAAATTTGGAGAGCCCTGATGGGGTCTGCGCTTATCTGGAATGCAAGCTTTAC 120  
 QY 1278 TGCCCTGGGCTCTCTGCGCTCATATCATGGAAGGAGGCT 1321  
 DB 121 TGCTCGATCTTCTCTGAGCTTACATCATGGAAGGAGGAGCT 164  
 RESULT 15  
 ABX47771  
 ID ABX47771 standard; cDNA; 397 BP.  
 AC ABX47771;  
 XX  
 DT 21-FEB-2003 (first entry)  
 XX  
 DE Bovine EST associated with lactation/muscle/fat deposition #12936.  
 XX  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM Gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 PN US2002137139-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-0960352.  
 XX  
 PR 12-JAN-1999; 99US-115707P.  
 PA 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 DR MPI; 2003-110599/10.  
 XX  
 PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID No 12936; 245bp; English.  
 CC The invention relates to a purified nucleic acid molecule associated with

CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
 CC  
 SQ Sequence 397 BP; 85 A; 115 C; 99 G; 98 T; 0 other;  
 Query Match 2.8%; Score 106; DB 25; Length 397;  
 Best Local Similarity 73.1%; Pred. No. 7.9e-22;  
 Matches 136; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1136 CTGTGCTCCAGGAGCAGCCAGGTATATGTTCTCTGACAAAGGAGGCTATACCTAGT 1195  
 DB 211 CTACCTCAGCCAGCGTACAGTATATGCTCTGACAAAGGAGGAGGCTATACCTCGT 270  
 QY 1196 AAGCGTTATCCGAAGCGGCTGGAAGAAGAGCGGAGCCCTCATGGGATATCTGGA 1255  
 DB 271 GCAGATATATCCAGCAAGCAAGCTGGAAGAGCGATTTGGAGCCCTGATGGGGTCTGCTTCA 330  
 QY 1256 CTCTGGAATGCGGCTTTATCTGCGCTGGGTCTTCTCGGCTCATATCATGAGAGTGA 1315  
 DB 331 TTCTGCGCTGCAAGCTTTATCTGCTGGAATCTTCTCAGCTCTACATCATGAGAGCCA 390  
 QY 1316 GGGGCT 1321  
 DB 391 GAAGCT 396

Search completed: December 15, 2003, 12:36:53  
 Job time : 592.508 secs



KW gene analysis; cattle breeding.  
 XX Bos Taurus.  
 OS US2002137139-A1.  
 PN 26-SEP-2002.  
 PD 24-SEP-2001; 2001US-0960352.  
 PF 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 PI WPI; 2003-110599/10.  
 DR  
 XX New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID No 6846; 245bp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
 CC  
 XX  
 SO Sequence 383 BP; 85 A; 108 C; 94 G; 96 T; 0 other;  
 Query Match 3.1%; Score 117.4; DB 25; Length 383;  
 Best Local Similarity 63.7%; Pred. No. 2.2e-25;  
 Matches 244; Conservative 0; Mismatches 56; Indels 83; Gaps 1;  
 QY 927 TTCCACAGGAGCCCACTACTGCGGTCTGACACACAGCGGAGTGGCGATAGCTGGCC 986  
 DB 84 TTTCAGTGAGAACCACTACTGCGGTCTGACACACAGCGGAGTGGCGATAGCTGGCT 143  
 QY 987 CATTGCTCATCTGCGCCCAAGGCTCTTTCAGACAGTGAATGCTCTTTTCTCTGGAGAA 1046  
 DB 144 CATTGAGCATCTGCGCCCAAGGCTCTTTCAGACAGTGAATGCTCTTTTCTCTGGAGTAA 203  
 QY 1047 AAAACCTCTATCTGCTCAAGGTGTGTATTTGGGGAGAGGCTTGAAGTGAAGACTGGACAA 1106  
 DB 204 AAGGCTCTAGC----- 214

QY 1107 GCATATCAACTCTGTATTATTATACATCTTTGCTCCAGGGACCCAGATATATGTC 1166  
 DB 215 -----TATCAGGGTACCCAGGATATATC 240  
 QY 1167 TTCTGACAAAGGAGGACTATACCTTAGTAAGCGGTTATCCGAGCGCTGAGAAAGAA 1226  
 DB 241 TTCTGACAAAGGAGGACTATACCTTAGTAAGGATTTATCCATAGCAGCTGAGAAAGAA 300  
 QY 1227 GTGAGGACCCCTCATGGAATTAATCTGACTGTGTGAATGCGGCTTTATCTCCCTGGG 1286  
 DB 301 TTGAGGAGCCCTGATGAGGTCTGCTTCATTTGTGTAAGACACCTTACTCTGCTGGA 360  
 QY 1287 TCTTCTGCGCTCATATCAGGC 1309  
 DB 361 TCTTCTGAGCTTACATCATTGC 383  
 RESULT 14  
 ABX43075  
 ID ABX43075 standard; cDNA; 228 BP.  
 AC ABX43075;  
 XX  
 DT 20-FEB-2003 (first entry)  
 XX  
 DE Bovine EST associated with lactation/muscle/fat deposition #8240.  
 XX  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.  
 XX  
 OS Bos Taurus.  
 XX  
 PN US2002137139-A1.  
 PD 26-SEP-2002.  
 PF 24-SEP-2001; 2001US-0960352.  
 PR 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 PI WPI; 2003-110599/10.  
 DR  
 XX New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID No 8240; 245bp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the

CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
XX  
SQ Sequence 420 BP, 90 A, 114 C, 116 G, 100 T, 0 other;

Query Match 3.3%; Score 123.6; DB 25; Length 420;  
Best Local Similarity 79.0%; Pred. No. 2.8e-27;  
Matches 147; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1136 CTTTGTCTCCAGGACCCAGGATATGCTTCTCTGACCAAGGAGGCTTATCCCTAGT 1195  
DB 93 CTACTCTATCAGAGGATCCAGGATATATCTTCTGACCAAGGACGCTTACCTCTGCT 152  
QY 1196 AAGCGGTTATCCGAGCGCTGAGAGAGAGTCCGAGCCCTCATGGAATTAATCTGGA 1255  
DB 153 AAAAGATTATCCAAAGAGCTGAGAGAGAAATTGGAGAGCTGATGGGCTTGCTTCA 212  
QY 1256 CTCTGTGATGCGGCTTTATCTGCTGCTGCTCTTCTGCTCCATATCATGCAAGTGA 1315  
DB 213 TTCTGTGAGTACGAGCTTTTACTGCTGCTGATCTTCTCACTCATGAGCAGCCA 272  
QY 1316 GGGGCT 1321  
DB 273 GAAGCT 278

RESULT 12  
ABX40203  
ID ABX40203 standard; cDNA; 374 BP.  
XX ABX40203;  
AC  
XX 20-FEB-2003 (first entry)  
DT  
XX  
XX Bovine EST associated with lactation/muscle/fat deposition #5368.  
DE  
XX  
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
KM muscle deposition; fat deposition; genome mapping; gene identification;  
XX gene analysis; cattle breeding.  
OS  
XX Bos Taurus.  
XX  
XX US2002137139-A1.  
XX  
XX 26-SEP-2002.  
PD  
XX  
XX 24-SEP-2001, 2001US-0960352.  
PF  
XX  
XX 12-JAN-1999; 99US-115707P.  
PR 11-JAN-2000; 2000US-0480902.  
XX  
XX  
XX (BYAT/) BYAT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARRE/) WARREN W C.  
XX  
XX  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
PI  
XX  
XX WPI, 2003-110599/10.  
DR  
XX  
XX New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and  
PT analysis, cattle breeding, or for genetically improving cattle  
XX  
XX Claim 2; SEQ ID No 5368; 245bp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMPD), derived

CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridization between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the level or pattern of the  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMPD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
XX  
SQ Sequence 374 BP, 86 A, 102 C, 92 G, 94 T, 0 other;

Query Match 3.2%; Score 118.2; DB 25; Length 374;  
Best Local Similarity 64.1%; Pred. No. 1.2e-25;  
Matches 243; Conservative 0; Mismatches 53; Indels 83; Gaps 1;

QY 927 TTCCACAGGAGACCACTACTGCGGTGACACACAGCCGGAGTGGCTGATGCTGGCC 986  
DB 79 TTTCAGTGAACACCTACTGCGGTGACACACAGCCGGAGTGGCTGATGCTGGCC 138  
QY 987 CATTCCTCATAGTGGAGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTCTGGAGAA 1046  
DB 139 CATTCAGTCTGTGGAGCCCGAGGCTCTTCAGCAGTGAATGCTCTTTCTCTGGAGAA 198  
QY 1047 AAAACTCTATCTGCTCAGGTGTGATTTGGGGAGAGGCTTGAAGTAGAGACTGGACAA 1106  
DB 199 AAGGCTCTAC----- 209  
QY 1107 GCATATCCAACTGTGATTTATTACATCTTTGCTCCAGGAGCAACCAAGATATATGTC 1166  
DB 210 -----TCATCAGAGGTATCCAGGATATATC 235  
QY 1167 TTCTGACAAAGAGAGGCTATACCTGATAGCGTTATCCAGAGCGCTGAGAGAA 1226  
DB 236 TTCTGACAAAGAGAGGCTATACCTGATAGCGTTATCCAGAGCGCTGAGAGAA 295  
QY 1227 GTCCGAGCCCTCATGAGATTAATCTGACCTGTGAGATGCGCCTTATATGCTGAG 1286  
DB 296 TTTCGAGGCTCATGAGGCTGCTGCTTCAATCTGTGAGATGAGCCTTATCTGCTGGA 355  
QY 1287 TCTTCTGAGCTCATATCA 1305  
DB 356 TCTTCTCAGCTCTACATCA 374

RESULT 13  
ABX41681  
ID ABX41681 standard; cDNA; 383 BP.  
XX ABX41681;  
AC  
XX 20-FEB-2003 (first entry)  
DT  
XX  
XX Bovine EST associated with lactation/muscle/fat deposition #6846.  
DE  
XX  
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
KM muscle deposition; fat deposition; genome mapping; gene identification;

QY 1288 CTTCTGCGCTCATATCATGAGGAGGCGCTTCTGGG 1327  
 DB 476 CTTCCAAGCTCTACGTACATCAGAGCGGCGTTGGTG 515

RESULT 10  
 ID ABT09465 standard; DNA; 689 BP.  
 AC ABT09465;  
 XX  
 DT 05-DEC-2002 (first entry)  
 XX  
 XX Phase-1 Rat CT gene SEQ ID No 553.  
 DE  
 XX Rat; toxicity study; rat toxic response gene; toxicological response;  
 KM drug development; phase-1 rat CT gene; ds.  
 XX  
 XX Ratus sp.  
 OS  
 XX MO20026682-A2.  
 PN  
 XX 29-AUG-2002.  
 PD  
 XX 29-JAN-2002; 2002MO-US02935.  
 XX  
 XX 29-JAN-2001; 2001US-264933P.  
 PR  
 XX 26-JUL-2001; 2001US-308161P.  
 XX  
 PA (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.  
 PI  
 XX Farlie G, Hicken SH, Farr SB;  
 XX  
 XX WPI; 2002-674961/72.  
 DR  
 XX  
 XX Evaluating the toxicity of an agent, useful in drug development or in  
 PT determining toxicological responses to a new drug, by determining the  
 PT expression of rat toxicologically relevant genes in the test animal in  
 PT response to the test agent -  
 XX  
 XX Disclosure; Page 241; 388pp; English.  
 PS  
 XX The invention relates to a method used for evaluating the toxicity of an  
 CC agent comprising determining the expression of a rat toxic response  
 CC gene(s) in the test animal in response to the agent. The method is useful  
 CC in drug development, particularly for conducting toxicity studies and  
 CC analysis before a new drug or compound is approved for human consumption  
 CC or use. The method is also useful in determining toxicological responses  
 CC to a new drug. This polynucleotide sequence represents a phase-1 rat CT  
 CC gene of the invention.  
 XX  
 XX Sequence 689 BP; 141 A; 185 C; 190 G; 172 T; 1 other;

Query Match 3.6%; Score 136; DB 24; Length 689;  
 Best Local Similarity 65.5%; Pred. No. 5.5e-31;  
 Matches 262; Conservative 0; Mismatches 55; Indels 83; Gaps 1;

QY 928 TCACAGGAGGACCACTACGTGCGCTGAGACACCGCGGATGCTGCACTAGCTGGCCC 987  
 DB 121 TTCAGTGGCTCCCACTACGTGCGCTGAGACACCGCGGATGCTGCACTAGCTGGCCC 180  
 QY 988 ATTGCTCATAGTGGCGCCGAGGCTCTTCAGAGTGAATGCTGCTTCTGGAGAGAA 1047  
 DB 181 ATTGCTCATAGTGGCGCCGAGGCTCTTCAGAGTGAATGCTGCTTCTGGAGAGAG 240  
 QY 1048 AAATCTATCTGCTGTCAGGTGTGTATTGGGGAGAGGCTTGAGGTAGAGACTGGAGCAAG 1107  
 DB 241 AAAGTCTATCTGA----- 253  
 QY 1108 CATATCAACTCTATATTATTATACATCCTTTGCTCCAGAGGACCCAGATATATGCT 1167  
 DB 254 -----TCAGAGGACCTCAAGATATATGCT 277

QY 1168 TCCTGACAAAGGAGGCTATACCTAGTAGAGCGTTATCCGAGCGGCTGAGAGAGAG 1227  
 DB 278 TCCTGACAAAGGAGGCGCAATTAACCTAGTAGAGCGTTATCCGAGCGGCTGAGAGAGAG 337

QY 1228 TCGGAGACCCCTCATGAGGAATTATCTGAGCTCTGTGATGCGGCTTTATCTGCTGGGT 1287  
 DB 338 TTGGAGAGCCCTCCCGGATCAGCTGTGATACCATATGATCAGCTTTCTGCTCGGTT 397

QY 1288 CTTCTGCGCTCATATCATGAGGAGGCGCTTCTGGG 1327  
 DB 398 CTTCCAAGCTCTACGTACATCAGAGCGGCGTTGGTG 437

RESULT 11  
 ID ABX45158 standard; cDNA; 420 BP.  
 XX  
 XX ABX45158;  
 AC  
 XX  
 XX 21-FEB-2003 (first entry)  
 DT  
 XX  
 XX Bovine EST associated with lactation/muscle/fat deposition #10323.  
 DE  
 XX  
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM gene analysis; cattle breeding.  
 XX  
 XX Bos Taurus.  
 OS  
 XX  
 XX US2002137139-A1.  
 PN  
 XX  
 XX 26-SEP-2002.  
 PD  
 XX  
 XX 24-SEP-2001; 2001US-0960352.  
 PF  
 XX  
 XX 12-JAN-1999; 99US-115707P.  
 PR  
 XX 11-JAN-2000; 2000US-0480902.  
 XX  
 XX (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 PI  
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX  
 XX WPI; 2003-110599/10.  
 DR  
 XX  
 XX New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle -  
 PT  
 XX  
 XX Claim 2, SEQ ID No 10323; 245bp; English.

The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 1512 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are: (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 1512 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome

PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-110599/10.  
XX  
XX  
XX New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and  
PT analysis, cattle breeding, or for genetically improving cattle  
XX  
XX  
XX Claim 2; SEQ ID No 7040; 245bp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMPD), derived  
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMPD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
XX  
XX  
SQ Sequence 408 BP; 95 A; 109 C; 111 G; 93 T; 0 other;  
Query Match 4.5%; Score 167; DB 25; Length 408;  
Best Local Similarity 79.8%; Pred. No. 9e-41;  
Matches 197; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 1479 CAGACGCGCGCTGTGTGCTGACCTGAAGTCAGAGGCCCAAGCAGGTGACAGAGC 1538  
DB 162 CAGGCGAAGAGCTGTGAGAGCTGACCTTAAGCTTGAAGAGCTCAAGCCAGTGAACAGAGC 221  
QY 1539 TTCCTTGGCCCCATGAGAGGTAGACGAGCCTGTGTATGAGAAAGTCCCTTGGCCCTTA 1598  
DB 222 TTCCTTGGCTCCATAGAGAAAGTGAAGGAGCCCTGTGTATGAGAGAGTCTCTGGGCCCTC 281  
QY 1599 ACTCATGTTCCGCAATGTCCTGCTTGTACTCATCATGCTGCTCCCAATTTGTACTGCT 1658  
DB 282 ACTCCGTTCTGCAATGTCCTGAGGCTGTGACTCTGCTCAAGGCCCAATCTTACTGCT 341  
QY 1659 ACAGATATGAGAGAAATGAAATGACAGCCCAAGGCCCTTCCGCAACCCCAAGATGTACCA 1718  
DB 342 ACAGAAATGAGAGAAATGAGCAAGACCAAGACCTTCCCAAGGCCCAAGAGATGAACA 401  
QY 1719 GTCTCTCT 1725  
DB 402 GCCTCTCT 408  
RESULT 9  
ABT08987  
ID ABT08987 standard; DNA; 631 BP.

XX  
XX ABT08987;  
AC  
XX  
XX  
DT 05-DEC-2002 (first entry)  
XX  
XX Phase-1 Rat CT gene SEQ ID No 75.  
DE  
XX  
XX Rat; toxicity study; rat toxic response gene; toxicological response;  
KM drug development; phase-1 rat CT gene; dr.  
XX  
XX Rattus sp.  
XX  
XX W0200266682-A2.  
PN  
XX  
XX 29-AUG-2002.  
PD  
XX  
XX 29-JAN-2002; 2002MO-US02935.  
PF  
XX  
XX 29-JAN-2001; 2001US-264933P.  
PR  
XX  
XX 26-JUL-2001; 2001US-308161P.  
PR  
XX  
XX (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.  
PA  
XX  
XX Parls G, Hicken SH, Farr SB;  
PI  
XX  
XX WPI; 2002-674961/72.  
DR  
XX  
XX  
XX  
XX Evaluating the toxicity of an agent, useful in drug development or in  
PT determining the toxicological responses to a new drug, by determining the  
PT expression of rat toxicologically relevant genes in the test animal in  
PT response to the test agent  
XX  
XX  
PS Disclosure; Page 124; 388bp; English.  
XX  
XX The invention relates to a method used for evaluating the toxicity of an  
CC agent comprising determining the expression of a rat toxic response  
CC gene(s) in the test animal in response to the agent. The method is useful  
CC in drug development, particularly for conducting toxicity studies and  
CC analysis before a new drug or compound is approved for human consumption  
CC or use. The method is also useful in determining toxicological responses  
CC to a new drug. This polynucleotide sequence represents a phase-1 rat CT  
CC gene of the invention.  
XX  
XX  
SQ Sequence 631 BP; 139 A; 168 C; 169 G; 154 T; 1 other;  
Query Match 3.6%; Score 136; DB 24; Length 631;  
Best Local Similarity 65.5%; Pred. No. 5.2e-31;  
Matches 262; Conservative 0; Mismatches 55; Indels 83; Gaps 1;  
QY 928 TCACAGGAGACCCACTACTGCGGCTGTGACACACAGCCGGAGTGGCATAGCTGGCCCC 987  
DB 199 TTGAGTGGCTCCCACTACTGCGGCTGTGACCTCAAGCCGATGAGGTGGCATAGCTGGCCCC 258  
QY 988 ATTGCTCATAGTGGCCCGAGGCTCTTCAGAGATGATGCTGCTTTCTCTGGAGAA 1047  
DB 259 ATTGCTCATAGTGGCCCGAGGCTCTTCAGAGATGATGCTGCTTTCTCTGGAGTGG 318  
QY 1048 AAATCTATCTGCTCAGGTGTGTATGAGGAGAGCCCTTGAAGTGAAGACTGGGACAAG 1107  
DB 319 AAAGTCTATCTGA----- 331  
QY 1108 CATATCAACTCTGTATTTATTATACATCTTTTCTCCAGGGGACCCAGTATATGCT 1167  
DB 332 -----TCCAGGGCACTCAAGTATATGCT 355  
QY 1168 TCTTGACAAAGGAGGCTATACCTAGTAAGCGGTATCCGAAGCGCTGAGAGGAAG 1227  
DB 356 TCTTGACAAAGGAGGCAATACCTAGTAAGCGGTATCCGAAGCGCTGAGAGGAAGC 415  
QY 1228 TCGGAGCCCTCATGAGGATTAATCTGAGACTGTGTGATCGGCTTTATCTGCCCTGGGT 1287  
DB 416 TTGGAGCCCTCCCGGATCAGCCTGTGAATACCATAGATGACACCTTTTCTGCCCTGGTT 475

QY 1479 CAGGACGCGGCTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGTGGACAGAGC 1538  
 DB 366 CAGGCCAAGAGCTGTGAGGCTGACCTGAAGTTCAGAGCTCAAGCCAGTGGACAGAGA 307  
 QY 1539 TTCCCTTGGCCCAATGAGAAAGTACGAGCCCTGTGTATGAAAGTCCCTTGGCCCTA 1598  
 DB 306 TTCTCTTGGCTCCATAGGAAGTGGAGGCGCTGTGTATGAGAAAGTATCTGGGCCCC 247  
 QY 1599 ACTCATGTTCCGCAATGCTCCGCTTGTACTCATCCATGGTCCCAATTTGTACTGCT 1658  
 DB 246 ACTCCGATGTGCAATGTCTGGGCTGTGACTCTGTCAGAGCCCAATCTGTACTGCT 187  
 QY 1659 ACAGATGATGAGAACTGAAATGCAAGCCAGGCGCTTCCGCAACCCAGAAATGTAGCA 1718  
 DB 186 ACAAGATGTGAGAAATGTGACACACCAAGAACTTTCGCAAGCCCAAGATGAAACA 127  
 QY 1719 GTCTCTGCGCTGCACTGACCTGAGGCGCTTGTGACATGATCTGGCTGGCCCACTC 1778  
 DB 126 GCTCTCGGGCTGCGCTCCCAACAGACCTCTGATGCGAGTTTGGCTCATGATCTCTC 67  
 QY 1779 CTAG-TTCTCATATTAAGACAGATTGCTTCTGCTT 1816  
 DB 66 CCAGTTTCATCAATTAAGCCAGAGTGTCTTCACT 28

RESULT 7  
 ABX46151  
 ID ABX46151 standard; cDNA; 387 BP.  
 AC ABX46151;  
 XX  
 DT 21-FEB-2003 (first entry)  
 DE Bovine EST associated with lactation/muscle/fat deposition #11316.  
 XX  
 KM Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 PN US2002137139-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-0960352.  
 XX  
 PR 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX  
 DR WPI; 2003-110599/10.  
 XX  
 PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID No 11316; 245bp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived  
 CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-

CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule, and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridization between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMFD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMFD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docID=20020137139.  
 XX  
 SQ Sequence 387 BP; 86 A; 108 C; 107 G; 86 T; 0 other;

Query Match 4.8%; Score 180.6; DB 25; Length 387;  
 Best Local Similarity 78.5%; Pred. No. 5.2e-45;  
 Matches 216; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1479 CAGGACGCGGCTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGTGGACAGAGC 1538  
 DB 111 CAGGCCAAGAGCTGTGAGGCTGACCTGAAGTTCAGAGCTCAAGCCAGTGGACAGAGA 170  
 QY 1539 TTCCCTTGGCCCAATGAGAAAGTACGAGCCCTGTGTATGAAAGTCCCTTGGCCCTA 1598  
 DB 171 TTCTCTTGGCTCCATAGGAAGTGGAGGCGCTGTGTATGAGAAAGTATCTGGGCCCC 230  
 QY 1599 ACTCATGTTCCGCAATGCTCCGCTTGTACTCATCCATGGTCCCAATTTGTACTGCT 1658  
 DB 231 ACTCCGATGTGCAATGTCTGGGCTGTGACTCTGTCAGAGCCCAATCTGTACTGCT 290  
 QY 1659 ACAGATGATGAGAACTGAAATGCAAGCCAGGCGCTTCCGCAACCCAGAAATGTAGCA 1718  
 DB 291 ACAAGATGTGAGAAATGTGACACCAAGAACTTTCGCAAGCCCAAGATGAAACA 350  
 QY 1719 GTCTCTGCGCTGCACTGACCTGAGGCGCTTGTGA 1753  
 DB 351 GCTCTCGGGCTGCGCTCCCAACAGACCTCTCTGA 385

RESULT 8  
 ABX41875  
 ID ABX41875 standard; cDNA; 408 BP.  
 AC ABX41875;  
 XX  
 DT 20-FEB-2003 (first entry)  
 DE Bovine EST associated with lactation/muscle/fat deposition #7040.  
 XX  
 KM Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 PN US2002137139-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-0960352.  
 XX  
 PR 12-JAN-1999; 99US-115707P.  
 PR 11-JAN-2000; 2000US-0480902.  
 XX

AAK53438;  
16-NOV-2001 (first entry)  
Murine transport and binding associated protein encoding cDNA SEQ. ID 3.  
Murine; liver; gene library; amino acid synthesis; binding protein;  
cell metabolism; energy metabolism; fatty acid metabolism; synthesis;  
phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;  
replication; transcription; translation; transport protein; ss.  
Mus musculus.  
DE20103510-U1.  
07-JUN-2001.  
28-FEB-2001; 2001DE-2003510.  
02-DEC-1999; 99DE-1058160.  
(LION-) LION BIOSCIENCE AG.  
WPI; 2001-368570/39.  
Gene library containing sequences with specific 3'-ends and no polyA  
tail, encoding proteins involved in a wide range of cellular processes  
Claim 15; Page 31; 251pp; German.  
This invention describes a novel gene library (A) comprises a gene  
sequence (or its part) encoding a protein involved in amino acid  
synthesis, cellular/energy metabolism, metabolism of  
fatty acids/phospholipids, synthesis or breakdown of  
purines/pyrimidines/nucleosides/nucleotides, DNA  
replication/transcription/translation, or is a transport/binding protein.  
(A) are produced that correspond to the 3'-end of mRNA but without the  
polyA tail. They can be prepared more efficiently and with less effort  
than conventional libraries. AAK53436-AAK54275 represent fragments of the  
gene library described in the method of the invention.  
Sequence 339 BP; 83 A; 81 C; 90 G; 85 T; 0 other;  
Query Match 5.5%; Score 204; DB 22; Length 339;  
Best Local Similarity 77.4%; Pred. No. 2.5e-52;  
Matches 263; Conservative 0; Mismatches 70; Indels 7; Gaps 1;  
1475 TCCCAAGAGCGGCGCTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGTGGACA 1534  
339 TCATCGAGAGCGGCGCTGTGTGCTGACCTGAAGTCAGAGCCCAAGCCAGTGGACA 280  
1535 GAGCTTCTTGGCCCATGAGAGGTAAGCGAGCCTTGTATGAAAAGTCCCTTGGC 1594  
279 GAGGTTCCGCGCCCATGAGAGGTAAGTGAAGGCGCCCTGTGTTGAGCAAGTCTTGGC 220  
1595 CCAATCATGTTCCCGCCCATGAGTGTCCGCTTGAATCTATTCATGATGCTCCCATTTGTAC 1654  
219 CCAACACATGTTCTTCATGATGTTCCAGCTTATTCATGATGAGGCCCAATTTGTAC 160  
1655 TCGTACAGATGATGAGAAAGTAAGTGAAGCAAGCCCTTCGCAACCCAGAAATG 1714  
159 TCGTATAGCAGATGATGAGAAAGTAAGTGAAGCAAGCCCTTCGCTCAAGCTTCAGAAAGTG 100  
1715 ACCAGTCTCTGGGCTGCACTCACTGAGGGGCTTGTGACATGATGCTGGCTGGGCCCA 1774  
99 AACAGATCCTTGGCTGCACTCACTGAGAAAGC-----CCTGATATGAGCCAGCCAC 47  
1775 CCGCTAGTTCCTCATATTAAGACAGATGCTTCTTCCG 1814  
46 TCTCCATTTTCATCCTAATTAAGCAGATGTTTCTTAC 7

RESULT 6  
ABX41415/C  
ID ABX41415 standard; cDNA; 400 BP.  
XX  
XX ABX41415;  
AC  
XX 20-FEB-2003 (first entry)  
DT  
XX Bovine EST associated with lactation/muscle/fat deposition #6580.  
DE  
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
KM muscle deposition; fat deposition; genome mapping; gene identification;  
KM gene analysis; cattle breeding.  
XX  
XX Bos Taurus.  
OS  
XX US2002137139-A1.  
PN  
XX 26-SEP-2002.  
PD  
XX 24-SEP-2001; 2001US-0960352.  
PR  
XX 12-JAN-1999; 99US-115707P.  
PR  
XX 11-JAN-2000; 2000US-0480902.  
XX  
XX (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
PI  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
DR WPI; 2003-110599/10.  
XX  
XX New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and  
PT analysis, cattle breeding, or for genetically improving cattle  
XX  
XX Claim 2; SEQ ID No 6580; 245bp; English.  
PS  
XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMPD), derived  
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 1512 nucleotide  
CC sequences, appearing as ABX4836-ABX4947, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 1512 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridization between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMPD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 1512 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=200203137139.  
XX  
XX Sequence 400 BP; 86 A; 106 C; 115 G; 93 T; 0 other;  
Query Match 5.3%; Score 197.4; DB 25; Length 400;  
Best Local Similarity 75.8%; Pred. No. 3.2e-50;  
Matches 257; Conservative 0; Mismatches 81; Indels 1; Gaps 1;

Db 292 ACTGATGTTCCGCAATGTCGCCGCTGTAACCTCATTCATG-CCCAATTTTACTGCT 234  
 Qy 1659 ACGTATGTTGGAAGAACTGATGCAAGCCCTTCCGCAACCCGAGATGTACCA 1718  
 Db 233 ANAGTATGTGGGAAACTGAATGCA-NCMAAGCCCTTCCGCAACCCGAGATGTACCA 175  
 Qy 1719 GTTCCGAGGCTGCACTGAGGAGGCTTTCGACATGATGCTGGGCTGGCCCACTTC 1778  
 Db 174 GTTCTGAGGCTGCACTGAGGAGGCTTTCGACATGATGCTGGGCTGGCCCACTTC 118  
 Qy 1779 CTAGTCTCATATATATAGACAGATGCTTCTTCTGCTTCTGCTGAGGAGGCTTTCGAC 1838  
 Db 117 CTAG-TCCTCATATATATAGACAGATGCTTCTT-CTTCTGCTGAG-GGGCTTTCGAC 61  
 Qy 1839 TGAGTCTGAGGCTGAGGAGGCTTTCGACATGATGCTGGGCTGGCCCACTTC 1898  
 Db 60 TGAGTCTGAGGCTGAGGAGGCTTTCGACATGATGCTGGGCTGGCCCACTTC 3  
 Qy 1899 TT 1900  
 Db 2 TT 1  
 RESULT 4  
 AB283572  
 ID AB283572 standard; cDNA; 690 BP.  
 AC AB283572;  
 XX  
 DT 14-MAY-2003 (first entry)  
 DE Toxicologically relevant human nucleotide sequence #731.  
 XX  
 OS Toxicologically relevant gene; toxicological response; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02003016500-A2.  
 PD 27-FEB-2003.  
 XX  
 PF 16-AUG-2002; 2002MO-US26514.  
 XX  
 PR 16-AUG-2001; 2001US-313080P.  
 PA (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.  
 XX  
 PI Neft RE, Dunn RT, Adkins K, Pickett GG, Kier LD, Schmeisler K;  
 PI Alen P;  
 XX  
 DR WPI; 2003-268322/26.  
 XX  
 PT Determining a toxicological response to an agent, useful for screening  
 PT of drugs, comprising comparing the expression profile of one or more  
 PT human toxic response genes to a reference gene expression profile  
 PT indicative of toxicity -  
 XX  
 PS Claim 1, Page 241, 455pp; English.  
 XX  
 CC The present invention describes a method (M1) for determining a  
 CC toxicological response to an agent, which comprises comparing the  
 CC expression profile of one or more human toxic response genes to a  
 CC reference gene expression profile indicative of toxicity, and so  
 CC determining the presence of a toxic response to the agent. Also  
 CC described: (1) an array comprising one or more polynucleotides selected  
 CC from the genes corresponding to the partial sequences given in AB282842  
 CC to AB284764, or their fragments of at least 20 nucleotides, or  
 CC homologues; and (2) determining if a gene putatively identified to be a  
 CC toxic response gene plays a role on toxic response pathways by  
 CC determining the expression profile of the gene after exposure of cells  
 CC or a human subject to a known toxic pharmaceutical or industrial agent,  
 CC comprising: (a) exposing cells to an agent or isolating cells from a  
 CC human subject who was exposed to an agent; (b) obtaining the test gene

CC expression profile for a putatively identified toxic response gene after  
 CC exposure to a known toxic pharmaceutical or industrial agent; and  
 CC (c) comparing the test profile to the expression profile of a gene with  
 CC a similar function or comparing the test profile to the expression  
 CC profile of that gene after exposure to other known toxic compounds. The  
 CC methods are useful for predicting and determining toxicological responses  
 CC on a cellular, organ or system level. The arrays comprising the human  
 CC genes are useful for toxicological screening of drugs, pharmaceutical  
 CC compounds and chemicals.  
 XX  
 SQ Sequence 690 BP; 147 A; 181 C; 196 G; 164 T; 2 other;  
 Query Match 6.2%; Score 232.6; DB 25; Length 690;  
 Best Local Similarity 66.4%; Pred. No. 5.4e-61;  
 Matches 508; Conservative 0; Mismatches 4; Indels 253; Gaps 2;  
 Qy 927 TTCCAGAGGAGCCCACTAGTGGGCTGTCGACACAGCCGGAGTGGCTGACATGCTGCC 986  
 Db 71 TTTCAGTGGAGCCCACTAGTGGGCTGTCGACACAGCCGGAGTGGCTGACATGCTGCC 130  
 Qy 987 CATGCTCATCAATGAGGAGGAGGCTTCCAGAGGAGTGGCTTCTTCCGAGGAGA 1046  
 Db 131 CATGCTCATCAATGAGGAGGAGGCTTCCAGAGGAGTGGCTTCTTCCGAGGAGA 190  
 Qy 1047 AAAACTCTATCTGTCAGAGTGTGATTTGGGGAGAGGCTTGAAGTGAAGACTGGAGCA 1106  
 Db 191 AAAACTCTATCTG----- 204  
 Qy 1107 GCATATCAACTCTGATTTATTATACCATCTTGTCTCCAGAGGAGCCAGATATATGTC 1166  
 Db 205 -----TCAGAGGAGCCAGGATATATGTC 227  
 Qy 1167 TTCTCGAAGAGGAGGCTTATACCTTATAGAGGCTTATCCAGAGGAGGCTGAGAGAGAA 1226  
 Db 228 TTCTCGAAGAGGAGGCTTATACCTTATAGAGGCTTATCCAGAGGAGGCTGAGAGAGAA 287  
 Qy 1227 GTGAGGAGCCCTCATGAGGATTAATCTGAGACTGTGAGATGCGGCTTTATCTGCGTGG 1286  
 Db 288 GTGAGGAGCCCTCATGAGGATTAATCTGAGACTGTGAGATGCGGCTTTATCTGCGTGG 347  
 Qy 1287 TCTTCTGAGCTTCATATCAATGAGGAGGCTTCTGAGGCTTATGAGAGGAGGCTTGT 1346  
 Db 348 TCTTCTGAGCTTCATATCAATG----- 369  
 Qy 1347 TCTGCTACTGTCTGTGAGCATATGATCCCAAGGAGGATGAGAGGCTTATGATGATC 1406  
 Db 370 ----- 369  
 Qy 1407 CCCAGGAGTGAAGAGGCTTATGATGATGATCCCATGATGAGAGGATGATGATGTTG 1466  
 Db 370 ----- 369  
 Qy 1467 GTGCTTCTCCAGAGAGGAGGCTGTGAGTGTGAGCTGAGCTGAGAGGAGGAGGAGGCA 1526  
 Db 370 -----CAGAGGAGGAGGCTGTGAGTGTGAGCTGAGAGGAGGAGGAGGAGGCA 417  
 Qy 1527 CGTGAAGAGGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1586  
 Db 418 CGTGAAGAGGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 477  
 Qy 1587 CCTTGGAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1646  
 Db 478 CCTTGGAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
 Qy 1647 ATTTGATGCTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1691  
 Db 538 ATTTGATGCTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582  
 RESULT 5  
 AAK53438/c  
 ID AAK53438 standard; cDNA; 339 BP.  
 XX



CC of the invention.

XX Sequence 1610 BP; 355 A; 439 C; 444 G; 372 T; 0 other;

Query Match 11.9%; Score 446; DB 22; Length 1610;  
 Best Local Similarity 73.7%; Pred. No. 4.8e-127;  
 Matches 722; Conservative 0; Mismatches 5; Indels 253; Gaps 2;

```

QY 928 TCACAGGAGGACCCACTACTGGGCTCTGACACACAGCGGGATGGCTGGATAGCTGGCCC 987
DB 880 TTCAGTGGAGCCACTACTGGGCTCTGACACACAGCGGGATGGCTGGATAGCTGGCCC 939
QY 988 ATTGCTCATAGTGGCCCGACAGGCTCTTCCAGCAGTGAATGCTCTTTCTTGGGAAGA 1047
DB 940 ATTGCTCATAGTGGCCCGACAGGCTCTTCCAGCAGTGAATGCTCTTTCTTGGGAAGA 999
QY 1048 AAACCTTATCTGTCAGAGTGTTATTTGGGGAGAGGCTTGAAGTGAAGA CTGGGACAAG 1107
DB 1000 AAACCTTATCTG----- 1012
QY 1108 CATATCAACTCTGTATTTATTACATCCTTTGTCTCCAGGGCAACCAAGTATATGCT 1167
DB 1013 -----TCCAGGGCAACCAAGTATATGCT 1036
QY 1168 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAGCGGCTGAGAGAGAG 1227
DB 1037 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAGCGGCTGAGAGAGAG 1096
QY 1228 TCGGAGCCCTCATAGGAGTATATCTGAGCTCTGAGATCGGGCTTTATCTGCGCTGGGT 1287
DB 1097 TCGGAGCCCTCATAGGAGTATATCTGAGCTCTGAGATCGGGCTTTATCTGCGCTGGGT 1156
QY 1288 CTTCGCGCTCATATCATAGTGAAGGAGGAGGCTTGGGTGCTTGAAGGAGGAGCTTTGT 1347
DB 1157 CTTCGCGCTCATATCATAG----- 1177
QY 1348 CTGCTACTGTCTGTGGCAGATAGATCCCAACAGAGGAGTGAAGAGGCTAGTCAAGATCC 1407
DB 1178 ----- 1177
QY 1408 CCAAGGCGATGAAGAGGCTAGGCTCAGAGATCCCATGATGAGAGGATGTAATGTTTGG 1467
DB 1178 ----- 1177
QY 1468 TGCGTTCTCCAGAGCGCGCTGTGGTGGCTGAGACCTGAAGTCAAGAGCCCAAGCCAC 1527
DB 1178 -----CAGAGCGCGGCTGTGGTGGCTGAGACCTGAAGTCAAGAGCCCAAGCCAC 1226
QY 1528 GTGACAGAGCTTCTTGGCCCATGAGAGGTAAGAGGAGCTTGTATGGAAGAAAGTC 1587
DB 1227 GTGACAGAGCTTCTTGGCCCATGAGAGGTAAGAGGAGCTTGTATGGAAGAAAGTC 1286
QY 1588 CTTTGGCCCTTAATCTCATATGTTCCGCAATGATGTCCTTGTATCTGATGCTCCAA 1647
DB 1287 CTTTGGCCCTTAATCTCATATGTTCCGCAATGATGTCCTTGTATCTGATGCTCCAA 1346
QY 1648 TTGTGACTGCTCAAGTATGTAAGAAACTGAATGAGCAAGGCCCTTCCGCAACCCA 1707
DB 1347 TTGTGACTGCTCAAGTATGTAAGAAACTGAATGAGCAAGGCCCTTCCGCAACCCA 1406
QY 1708 GAATGACACAGTCTCTGGGCTGACATCAGAGGGGCTTCTGACATGAGTCTGGCCT 1767
DB 1407 GAATGACACAGTCTCTGGGCTGACATCAGAGGGGCTTCTGACATGAGTCTGGCCT 1466
QY 1768 GGGCCCACTCTAGTCTCTCATATATMAAGACAGATGCTTCTGCTGCTTCTGAGAGG 1827
DB 1467 GGGCCCACTCTAGTCTCTCATATATMAAGACAGATGCTTCTGCTGCTTCTGAGAGG 1526
QY 1828 GCGTTCTGACATGAGTCTGGCGGCGCCCACTCCCAAGTTTCTCATATMAAGACAGAT 1887
DB 1527 GCGTTCTGACATGAGTCTGGCGGCGCCCACTCCCAAGTTTCTCATATMAAGACAGAT 1586
QY 1888 TGCTTCTTCACTTGAATCAA 1907

```

DB 1587 TGCTTCTTCACTTGAAGAAA 1606

RESULT 3  
 ABN96669/c  
 ID ABN96669 standard; DNA; 504 BP.

XX ABN96669;

DT 13-AUG-2002 (first entry)

DE Gene #3167 used to diagnose liver cancer.

KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

(GENE-) GENE LOGIC INC.

Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

WPI; 2002-426119/45.

DR Diagnosing and detecting the progression of liver cancer,  
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
 PT involves detecting the level of expression of two or more genes in a  
 PT liver tissue sample -

PS Claim 1; SEQ ID NO 3167; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumor in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytoplastic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 504 BP; 123 A; 119 C; 138 G; 110 T; 14 other;

Query Match 7.1%; Score 265.2; DB 24; Length 504;  
 Best Local Similarity 93.1%; Pred. No. 3.2e-71;  
 Matches 393; Conservative 0; Mismatches 16; Indels 13; Gaps 11;

```

QY 1480 AGAGCGGCGGCTGTGGTGGCTGAGACCTGAAGTCAAGAGCCCAAGCAGTGAAGAGCT 1539
DB 410 AGAGCGGCTGTGGTGGCTGAGACCTGAAGTCAAGAGCCCAAGCAGTGAAGAGAGCT 352
QY 1540 TCCCTTGG-CCCATGAAGAGTGAAGCGAGGCTTGTATGGAAGAAAGTCCCTTGGCCCTA 1598
DB 351 TCCCTTGGNCCCATGAAGAGTGAAGCGAGGCTTGTATGGAAGAAAG-NCCTTGGCCCTA 293
QY 1599 ACTGATGTTCCGCAATGATGTCCTGGGCTTGTATCTGATCAGATGATGCTCCCAATTGATGCT 1658

```



PT New cell and tissue specific polynucleotides useful for diagnosis,  
 PT prognosis or monitoring of treatments for disorders where the gene is  
 PT associated with a cancer, immunopathology or neuropathology -

PS Claim 1, Page 244; 327pp; English.

CC AAH5161 to AAH5756 represent cell and tissue specific polynucleotide  
 CC sequences (I). (I) can have cytostatic, immunomodulatory and  
 CC neuroprotective activities, and can be used in gene therapy. (I) and  
 CC proteins (II) encoded by them are used in high throughput screening  
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
 CC their fragments, immunoglobulins, inhibitors, drug compounds and  
 CC pharmaceutical agents. Expression of (I) in a sample indicates the  
 CC differentiation of embryonic stem cells into a tissue selected from  
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
 CC tissues. (I) and (II) are used to produce an expression profile that  
 CC defines a metabolic or developmental process, treatment, condition,  
 CC disease or disorder. The gene profile can be used for diagnosis,  
 CC prognosis or monitoring of treatments and for investigating a  
 CC predisposition to a disorder where the gene is associated with a  
 CC cancer, immunopathology or neuropathology.

CC Sequence 1571 BP; 344 A; 431 C; 434 G; 362 T; 0 other;

Query Match 11.9%; Score 446; DB 22; Length 1571;  
 Best Local Similarity 73.7%; Pred. No. 4.7e-127;  
 Matches 722; Conservative 0; Mismatches 5; Indels 253; Gaps 2;

QY 928 TCCAGAGGAGCCACTACTGCGCTGAGACACAGCCGGAGTGGCTGACCTGAGCC 987  
 DB 840 TTGAGTGGAGCCACTACTGCGCTGAGACACAGCCGGAGTGGCTGAGCC 899  
 QY 988 ATTGCTCATGATGGGCCCCAGGGTCTTCAGACAGTGGATGCTCTTTCTGGAGAA 1047  
 DB 900 ATTGCTCATGATGGGCCCCAGGGTCTTCAGACAGTGGATGCTCTTTCTGGAGAA 959  
 QY 1048 AACTCTATGTGTCAGAGTGTTATTTGGGGAGAGGCTTGAAGTGAAGCTGGACAG 1107  
 DB 960 AACTCTATGTGTCAGAGTGTTATTTGGGGAGAGGCTTGAAGTGAAGCTGGACAG 972  
 QY 1108 CATATCAACTGTGATTTATTTATACCATCTTGTCTCCAGGGACCCAGGATATGTCT 1167  
 DB 973 -----TCCAGGGACCCAGGATATGTCT 996  
 QY 1168 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAAGCGCTGAGAAAG 1227  
 DB 997 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAAGCGCTGAGAAAG 1056  
 QY 1228 TCGGAGACCCCTCATGAGATTTATCTGACCTCTGTGATGGGCTTTATTCGCTGGGT 1287  
 DB 1057 TCGGAGACCCCTCATGAGATTTATCTGACCTCTGTGATGGGCTTTATTCGCTGGGT 1116  
 QY 1288 CTCTCGGCTCCATATCATGATGAGGAGGCTTCTGGGTGTAGAGGGCAGCTTGT 1347  
 DB 1117 CTCTCGGCTCCATATCATG----- 1137  
 QY 1348 CTGCTACTGTCTGTGGCATATGATCCCAACAGGGCATGAGAGCGCTAGTCAGATCC 1407  
 DB 1138 ----- 1137  
 QY 1408 CAGGGCATGAGAGCGCTAGTCAGATCCCAACAGGGCATGAGAGCGCTATGTTGG 1467  
 DB 1138 ----- 1137  
 QY 1468 TGCCTTCTCCAGAGCGGCGCTGTGTGGCTGAGACCTTAGTCAAGAGCCCAAGCCAC 1527  
 DB 1138 -----CAGGAGCGGCGCTGTGTGGCTGAGACCTTAGTCAAGAGCCCAAGCCAC 1186  
 QY 1528 GTGAGACAGACTTCTTGGCCCCCATGAGAGGATGAGAGGAGCTTGTGTATGAGAAAGTC 1587  
 DB 1187 GTGAGACAGACTTCTTGGCCCCCATGAGAGGATGAGAGGAGCTTGTGTATGAGAAAGTC 1246

QY 1588 CCTGGCCCTAATCATATGTTCCGCAATGATCCGGCTTGTACTCATGATGATGCCAA 1647  
 DB 1247 CCTGGCCCTAATCATATGTTCCGCAATGATCCGGCTTGTACTCATGATGATGCCAA 1306  
 QY 1648 TTTGTAAGTCTAAGTATGATGAGAGAACTGATGAGCAAGGCCCTTCGCAACCCA 1707  
 DB 1307 TTTGTAAGTCTAAGTATGATGAGAGAACTGATGAGCAAGGCCCTTCGCAACCCA 1366  
 QY 1708 GAATGAGACAGTCTCTGGGCTGCACTCATGAGAGGCTTCTGACATGATGATGCCCT 1767  
 DB 1367 GAATGAGACAGTCTCTGGGCTGCACTCATGAGAGGCTTCTGACATGATGATGCCCT 1426  
 QY 1768 GGGCCCACTCTAGTCTCTAATTAATTAAGACAGATTTCTGCTCTGCACTGAGGG 1827  
 DB 1427 GGGCCCACTCTAGTCTCTAATTAATTAAGACAGATTTCTGCTCTGCACTGAGGG 1486  
 QY 1828 GCTTCTGACATGATGATGCTGGCTGGCCCACTCCCACTTTCTCATTAATTAAGACAGAT 1887  
 DB 1487 GCTTCTGACATGATGATGCTGGCTGGCCCACTCCCACTTTCTCATTAATTAAGACAGAT 1546  
 QY 1888 TGCCTTCTCATTTGATCA 1907  
 DB 1547 TGCCTTCTCATTTGATCAAAAA 1566

RESULT 2  
 AAH98592  
 ID AAH98592 standard; cDNA, 1610 BP.

XX AAH98592;  
 AC  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Human EST-derived coding sequence SEQ ID NO: 449.  
 XX  
 KM Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KM diagnostics; forensic test; gene mapping; genetic disorder;  
 KM biodiversity; gene therapy; nutrition; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200154477-A2.  
 PD  
 XX 02-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001WO-US02687.  
 XX  
 PR 25-JAN-2000; 2000US-0491404.  
 PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 XX  
 PA (HYSB-) HYSBQ INC.  
 XX  
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,  
 PI Cao Y, Drmanac RA, Zhang J, Weirman T;  
 XX  
 DR WPI; 2001-476164/51.  
 DR P-PSDB; AAM23933.  
 XX  
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 PS Claim 1, Page 498; 1275pp; English.  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 06:38:47 ; Search time 566.508 seconds  
(without alignments)  
17159.770 Million cell updates/sec

Title: US-09-900-448-3\_COPY\_10001\_13737

Perfect score: 3737  
Sequence: 1 cccaaggagagatcagtcac.....accttttgctgtctcaca 3737

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_15Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	446	11.9	1571	22	AAH57485 Human liver cell s
2	446	11.9	1571	22	AAH57485 Human liver cell s
3	265.2	7.1	504	24	ABN96659 Human EST-derived
4	232.6	6.2	690	25	ABN96659 Human EST-derived
5	204	5.5	339	22	AAK53438 Toxicologically re
6	197.4	5.3	400	25	ABX41415 Murine transport a
7	180.6	4.8	387	25	ABX46151 Bovine EST associa
8	167	4.5	408	25	ABX41875 Bovine EST associa

9	136	3.6	631	24	ABT08987
10	136	3.6	689	24	ABT09465
11	123.6	3.3	420	25	ABX45158
12	118.2	3.2	374	25	ABX40203
13	117.4	3.1	363	25	ABX41681
14	111.2	3.0	228	25	ABX43075
15	106	2.8	397	25	ABX47771
16	106	2.8	700	22	AAH92374
17	105.4	2.8	397658	25	ABO83210
18	103.2	2.8	203	25	ABX44025
19	102.2	2.7	343	25	ABX47807
20	101.8	2.7	123219	23	AAH88703
21	101.2	2.7	2674	22	AAH14203
22	101.2	2.7	183999	22	AAH92831
23	100.2	2.7	593	22	ABX59483
24	100.2	2.7	593	22	AAK07755
25	100.2	2.7	593	22	AAK33622
26	100.2	2.7	593	22	AAH39344
27	100.2	2.7	593	23	ABX33439
28	100	2.7	395	21	AACT10672
29	99.2	2.7	9160	24	ABK69800
30	97.6	2.6	9266	24	ABK69899
31	97.6	2.6	9268	24	ABK69898
32	97.6	2.6	12036	22	AAK69341
33	97.4	2.6	1578	24	ABX02154
34	96	2.6	12049	22	AAK69340
35	95.8	2.6	159095	25	ABZ80818
36	95.2	2.5	74822	24	ABT10752
37	95.2	2.5	397658	25	ABO83210
38	94.2	2.5	568	22	ABX63861
39	94.2	2.5	568	22	ABX31041
40	94.2	2.5	568	22	AAK12371
41	94.2	2.5	568	22	AAK38088
42	94.2	2.5	568	22	AAI18862
43	94.2	2.5	568	22	AAI43984
44	94.2	2.5	568	23	ABX37705
45	94.2	2.5	568	24	ABX12100

#### ALIGNMENTS

RESULT 1  
AAH57485  
ID AAH57485 standard; cDNA; 1571 BP.  
XX  
XX AAH57485;  
AC  
XX  
XX 10-SEP-2001 (first entry)  
DT  
XX  
XX Human liver cell specific cDNA sequence SEQ ID NO:325.  
DE  
XX  
XX Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
KW lung; liver; uterus; ovary; stomach; kidney; pancreas; se;  
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200132927-A2.  
PN  
XX  
XX 10-MAY-2001.  
PD  
XX  
XX 02-NOV-2000; 2000MO-US30396.  
PF  
XX  
XX 04-NOV-1999; 99US-0163508.  
PR  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
PA  
XX  
XX Sormasbe T, Seilhamer JJ, Watson GA;  
PI  
XX  
XX WPI, 2001-291057/30.  
DR  
XX

**THIS PAGE BLANK (USPTO)**

TITLE Direct Submission  
JOURNAL Submitted (14-SEP-1989) Tatum F.M., National Animal Disease Centre,  
P O Box 70, Ames IA 50010, U S A  
REFERENCE 2 (bases 1 to 1486)  
AUTHORS Morgan,W.T., Muster,P., Tatum,F., Kao,S.M., Alam,J. and Smith,A.  
TITLE Identification of the histidine residues of hemopexin that  
coordinate with heme-iron and of a receptor-binding region  
JOURNAL J. Biol. Chem. 268 (9), 6256-6262 (1993)  
MEDLINE 93203213  
PUBMED 7681064

FEATURES  
source Location/Qualifiers

1..1486  
/organism="Oryctolagus cuniculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9986"  
/cbase\_type="liver"  
/clone\_lib="lambda gcl1"  
3..1385  
/codon\_start=1  
/product="hemopexin"  
/protein\_id="CAA34452.1"  
/db\_xref="GI:433687"  
/db\_xref="SWISS-PROT:P20058"  
/translation="MYKASGIPALGVWGLCWSLATVNSVPLTSAHGVTEGSGTKP  
EADVIEQCSDFDATTLDNGMTLFPDSEPMKSHRGIRLISRMKNFICPVDA  
FRRHSVYLLIKGDYVYTSKNEKVPKSLDDEPFGIPEPLDAVECHRGCODEG  
ILFPGNRKMFMDLTGTGKKRSMKAVGNCSTSLRWLGRIYCPGNOPLRPNVSGEV  
PFGYPLVDYDPIFSLCPGRHRSRHSRSTQGHSTRCDDPLVLSANVSDHQAITYFS  
GSHYWRLDTRNDGWSHWPILAHQWPGQSTVDAAFSWEDKLYLIQDTKYVFFLYGGYT  
LVNGYPRLEKEKSPPIVLSLEAVDAFVCPGSSRLHMAGRLLMWLDKSGAQTMT  
ELPWPHEKYDGAICMEKPLGPNKSTSGPNLYIHGPNLYCYRHVDKLANAKNLPQ  
RVSRLIGCTH"

sig\_peptide  
mat\_peptide

polya\_signal 3..80  
BASE COUNT 354 a 381 c 393 g 358 t  
ORIGIN

Query Match 6.8%; Score 255.6; DB 4; Length 1486;  
Best Local Similarity 86.5%; Pred. No. 1.5e-68;  
Matches 294; Conservative 0; Mismatches 44; Indels 2; Gaps 1;

QY	1479	CAGGACGGCGGCTGTGTGCTGACCTGAAGTCAGAGGCCCAAGCCAGGTGACAGAGC	1538
DB	1123	CAGGACGGCGCTGTGTGCTGACCTGAAGTCAGAGCTCAAGCCAGGTGACAGAAC	1182
QY	1539	TTCTTGCCGCCATGAGAGGTAGACGAGCCTTGTGTATGAAAAATCCCTTGCCCTTA	1598
DB	1183	TTCTTGCGCTCATGAGAGGTGACGGGCGCTGTGTATGAAAAAGCCCTTGCGCCCA	1242
QY	1599	ACTCATGTTCCGCGCATGTGCTCCGGCTTGTACTCATTCATGATGATCCCAATTGTACTGCT	1658
DB	1243	ACTCATGTTCTACAGTGTGCTCCCACTTACTCATTCATGATGATCCCAATTGTATGCT	1302
QY	1659	ACAGTATGTGAGGAAGTGAATGACCAAGGCCCTTCGCAACCCCAAGATGTACCA	1718
DB	1303	ACAGGATGTGACCAACTCAATGACAGCCCAAGAACTCCCCAGCCCAAGAGTGTCCA	1362
QY	1719	GTCTCTGGGCTGACACTGAGGGGCTTCTGACATGATGATGCTGAGCTGACCCACCTC	1778
DB	1363	GACTCTGGGCTGTACTGACATGAGTGTCTCCGACATGATGATGAGCTGCTCTC	1422
QY	1779	C--TAGTTCTCTCAATAAAGACAGATGCTTCTTGCTT	1816
DB	1423	CTGATTTCTCTCAATAAAGCAGATTTCTTCTTCACTT	1462

Search completed: December 15, 2003, 23:40:52  
Job time : 8797.02 secs



LOCUS	HSHEPEXR	1485 bp	mRNA	linear	PRI 06-JUL-1988
DEFINITION	Human mRNA for serum beta-glycoprotein hemopexin.				
ACCESSION	X02537				
VERSION	X02537.1	GI:32061			
KEYWORDS	glycoprotein; hemopexin; serum glycoprotein.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1485)				
AUTHORS	Altiruda, F., Poli, V., Restagno, G., Argos, P., Cortese, R. and Silengo, L.				
TITLE	The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal, repeating homology				
JOURNAL	Nucleic Acids Res. 13 (11), 3841-3859 (1985)				
MEDLINE	85242073				
PUBMED	2989777				
COMMENT	Data kindly reviewed (07-JAN-1986) by L. Silengo.				
FEATURES	Location/Qualifiers				
source	1. 1485				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	7. 1323				
	/note="hemopexin; Protein sequence is in conflict with the conceptual translation"				
	/codon_start=1				
	/protein_id="CAA26382.1"				
	/db_xref="GI:1335098"				
	/db_xref="SWISS-PROT:P02790"				
	/translation="TLPPTSAHGVAAGETKPPDYTERCSDDGSPDATTLDNGTMTLFEFGFVWKSXKMDRLISERMAHNPSPVNAARQGNISVFLIKDVMVYPERKKEKGYKLIQDERPGIPSPLEDAVECHRGECQAEKGLFPGDNEWMDATGMKRSNPAVAGCSSALRLWLGRIYFCQGNQFLRPDVRREVPKRYPRDYRDFMPCPGHGHRNGTGHNSNTHGPEYMKCSPHLYLSALTSPNHGATYAFSGTHYRLDTSDDGHSWPIAHQWPGSPAVDAAPAFMRBEKLYLVQGTQVYVFLTKGTYLVSGYPRKLEKEVTPHGIILIDSVDAAFICPSSRLHIMAGRLMMWLDKSGAQAQWTELPMPEHKVDALCMKSLSGHNSCANGPGLYHIGPVLVCSDVPEKNNAAALPQPNVTSLLGCTH"				
	1375. 1380				
	/note="pot. polyA signal"				
	1459. 1464				
	/note="pot. polyA signal"				
	1485				
	/note="put. polyA site"				
	323 a 412 c 406 g 344 t				
BASE COUNT	323 a 412 c 406 g 344 t				
ORIGIN					
	11.5%; Score 431.2; DB 9; Length 1485;				
	Best Local Similarity 73.6%; Pred. No. 3.8e-124;				
	Matches 717; Conservative 0; Mismatches 3; Indels 254; Gaps 3;				
QY	928	TTCACAGGAGCCCACTACTGCGCGTCTGGACACACGCGGAGATGCTGGCAATAGCTGCGCC	987		
Db	766	TTCACTGGAGCCCACTACTGCGCGTCTGGACACACGCGGAGATGCTGGCAATAGCTGCGCC	825		
QY	988	ATTGCTCATCATGAGGCGCCAGAGGTCCTTCAACAGTGAATGCTGCTTTCCTGGGAGAGA	1047		
Db	826	ATTGCTCATCATGAGGCGCCAGAGGTCCTTCAACAGTGAATGCTGCTTTCCTGGGAGAGA	885		
QY	1048	AAACTCTATCTGATTCACAGTGTGTATTGGGAGAGAGCTTGAGGTAGAGACTGGGACAAAG	1107		
Db	886	AAACTCTATCTG-----	898		
QY	1108	CATATCCAACTCTGATTATTATACATCTCTTGTCTTCTCAAGGACCCAGGTATATGTCT	1167		
Db	899	-----TCCAGGAGACCCAGGTATATGTCT	922		
QY	1168	TCCTGACAAAGGAGGAGCTATACCTCACTAAGCGGTTATCCGAAGCGCTGAGAAAGAG	1227		
Db	923	TCCTGACAAAGGAGGAGCTATACCTCACTAAGCGGTTATCCGAAGCGCTGAGAAAGAG	982		

Oy		1228	TGGGACCCCTCAGTGGGAATTATCTTGAACTCTGTGANGGGCCTTTATCTGCCTGGGT	1287
Db		983	TGGGACCCCTCATGGGATTAATCTTGAACTCTGTGANGGGCCTTTATCTGCCTGGGT	1042
Oy		1288	CTTCTCGGCTTCATATCATGCGAAGTGAAGGGCTTCTGGGTCTTAAAGGCCACTTGT	1347
Db		1043	CTTCTCGGCTTCATATCATGCG-----	1063
Oy		1348	CTGTACTCTGTCTGTGGCATATGATCCCAACAGGGCATGAGAAGCCTAGTCAGATCC	1407
Db		1064	-----	1063
Oy		1408	CCAAGGCATGAGAAGGCTTAGTTCAGAGATCCCAATGACAATGAGCCATGCTATGTTTTG	1467
Db		1064	-----	1063
Oy		1468	TGCTTTCTCCCCAGAGACGGGGGCTGTGTGTGGCTTGAGACTTGAGAGGCCAACCCAC	1527
Db		1064	----- --CAGGACGGCGGGCTGTGTGTGGCTTGAGACTTGAGAGGCCAACCCAC	1112
Oy		1528	GTTGACAGAGCTTCCTTGGGCCCATGGAAGGTGAGACGAGGCCCTGTGTATGAAAAAGTC	1587
Db		1113	GTGACAGAGCTTCCTTGGGCCCATGGAAGGTGAGACGAGGCCCTGTGTATGAAAAAGTC	1172
Oy		1588	CCTTGGCCCTTAACATCATGTTCCGCGCAATGSGTCCGCGCTTGACTCATATCATGTGCCAA	1647
Db		1173	CTTGGCCCTTAACATCATGTTCCGCGCAATGSGTCCGCGCTTGACTCATATCATGTGCCAA	1232
Oy		1648	TTTGTACTGCTACAGTATGTGAGAAACTGAATGACGCCAAGGCCCTTCCGCAACCACA	1707
Db		1233	TTTGTACTGCTACAGTATGTGAGAAACTGAATGACGCCAAGGCCCTTCCGCAACCACA	1232
Oy		1708	GAATGTACCAAGTCTCCCTGGGGCTGACTCATCTGAGGGGGCTTGTGACATGAGTCTGGCT	1767
Db		1293	GAATGTACCAAGTCTCCCTGGGGCTGACTCATCTGAGGGGGCTTGTGACATGAGTCT--GCTT	1351
Oy		1768	GGCCCCCACTCTAGTTCCTCATATAATAAGACAGATTGCTTCTTGCTTCACTGAGGG	1827
Db		1352	GGCCCCCACTCTAGTTCCTCATATAATAAGACAGATTGCTTCTTGCTTCACTGAGGG	1411
Oy		1828	GCTTCTGACATGAGTCTGGGCTTGGGCCCACTTCCCAGATTTCTCATATAATAAGACAGAT	1887
Db		1412	GCTTCTGACATGAGTCTGGGCTTGGGCCCACTTCCCAGATTTCTCATATAATAAGACAGAT	1471
Oy		1888	TGCTTCTTCACTTG 1901	
Db		1472	TGCTTCTTCACTTG 1485	
<hr/>				
RESULT 13				
LOCUS	G11396/c	400 bp	DNA	linear STS 30-MAR-2000
DEFINITION	SHGC-10293 Human Homo sapiens STS genomic, sequence tagged site.			
ACCESSION	G11396			
VERSION	G11396.1 GI:1017488			
KEYWORDS	STS.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	1. (bases 1 to 400)			
TITLE	Olivier,M. and Cox,D.R.			
JOURNAL	Unpublished, Olivier, M., Cox, D.R. (2000)			
COMMENT	Unpublished (2000)			
<hr/>				
Contact: Michael Olivier, David R. Cox Stanford Human Genome Center Stanford University School of Medicine 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA Tel: (650) 320-5800 Fax: (650) 320-5801 Email: oliviere@hgsc.stanford.edu				

Db 1458 GGGCCCACTCCAGTTCCTCATATAAAGACAGATGCTTTCCTTCCTCACTAGGG 1517  
QY 1828 GCCTTCGACATAGTCTGGCCCTGGCCCACTCCCACTTCTCATATAAAGACAGAT 1887  
Db 1518 GCCTTCGACATAGTCTGGCCCTGGCCCACTCCCACTTCTCATATAAAGACAGAT 1577  
QY 1888 TGGCTTCCTCACTTGATCA 1907  
Db 1578 TGGCTTCCTCACTTGATCA 1597  
RESULT 11  
HUMHXA 1546 bp mRNA linear PRI 08-NOV-1994  
LOCUS Huma hemopexin mRNA, 3' end.  
DEFINITION J03048.1 GI:184487  
ACCESSION J03048.1 GI:184487  
KEYWORDS hemopexin.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Law, M.L., Cai, G.Y., Hartz, J.A., Jones, C. and Kao, F.T.  
TITLE The hemopexin gene maps to the same location as the beta-globin  
gene cluster on human chromosome 11  
JOURNAL Genomics 3 (1), 48-52 (1988)  
MEDLINE 89122012  
PUBMED 3220477  
COMMENT Original source text: Human, cDNA to mRNA, clone lambda-Hx5.  
Draft entry and printed copy of sequence for (Genomics (1988) In  
press) kindly provided by  
G.-Y. Cai and M.L. Law, 01-June-1988.  
FEATURES  
source  
1.1546  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="11p15.5-p15.4"  
1.1546  
/gene="HPX"  
/gene="HPX"  
/product="HxM mRNA"  
<1.1386  
/gene="HPX"  
/note="hemopexin precursor"  
/codon\_start=1  
/protein\_id="AA52704.1"  
/db\_xref="GI:386789"  
/translation="ARVLGAPVALGLMSLWSLAITPLPPTSAHGNVABGETKPD  
VTERCDGWSFDATLIDNGTMLFKGBFVWKSHKNDRELISRMKNFSPVDAARQ  
GHNSVPLIKGDKVVYPPEKKEKGYPLIQDEPGIPLDAVECHRGSCQEGVLE  
FOGDRWFMDLATGTWKERSMPAVGNCSSALRWLGRYVCGOQPLRPDVEBPVR  
YPRDVEDYFMPGCGRGHGRNGHNGSTHSGEYVNRCPHIVL.SALTSNDNGATAF  
SCNHYWR.LDTSBNGHMSWPIAHOWPGSPSVDAAFSEWBEKLYIVQSTQYVETTKGY  
TLVSGYPKRLKEKVGTPHGIILDSVDAAFICPSSRLHITAGRLWLMDKSGAQWTW  
TELPMFHEKVDGALCMKESLSPNSCSANGGLYLIGHPNLYCYSDVEKLNAAKALPOP  
QNTSLAGCTH"  
1.66  
/gene="HPX"  
/note="hemopexin signal peptide"  
67.1383  
/gene="HPX"  
/product="hemopexin"  
1380.1462  
/note="tandem repeat copy A"  
1463.1546  
/note="tandem repeat copy B"  
repeat\_region 330 a 427 c 430 g 359 t  
BASE COUNT 330 a 427 c 430 g 359 t  
ORIGIN Chromosome 11pter-p11.

Query Match 11.9%; Score 443.2; DB 9; Length 1546;  
Best Local Similarity 73.7%; Pred. No. 66-128;  
Matches 718; Conservative 0; Mismatches 3; Indels 253; Gaps 2;  
QY 928 TCACAGGAGACCCACTACTAGGCGTCTGACACCAAGCCGAGATGGCTGACATAGCTGGCCC 987  
Db 826 TTCAGTGGAGACCCACTACTAGGCGTCTGACACCAAGCCGAGATGGCTGACATAGCTGGCCC 885  
QY 988 ATTGCTCATCATAGGCGCCAGAGGTCTTTCAGACATGAGATGCTTTCCTGGGAGAA 1047  
Db 886 ATTGCTCATCATAGGCGCCAGAGGTCTTTCAGACATGAGATGCTTTCCTGGGAGAA 945  
QY 1048 AAATCTATCTGGTTCAGGTGTGTATTTGGGGAGAGGCTTGAAGTGAAGACTGGACAAG 1107  
Db 946 AAATCTATCTGG-----TTCAGGGCACCCAGGTATATGCT 958  
QY 1108 CATATCAACTCTGTATTTATTAACATCCTTGTCTCCAGAGGACACAGATATATGCT 1167  
Db 959 -----TTCAGGGCACCCAGGTATATGCT 982  
QY 1168 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAGCGGCTGGAGAGAG 1227  
Db 983 TCCTGACAAAGGAGGCTATACCTTAGTAAGCGGTTATCCGAAGCGGCTGGAGAGAG 1042  
QY 1228 TGGGAGCCCTCATGGGATTTATCTGAGCTGTGATGAGTGGGCTTTATCTGCGCTGGGT 1287  
Db 1043 TGGGAGCCCTCATGGGATTTATCTGAGCTGTGATGAGTGGGCTTTATCTGCGCTGGGT 1102  
QY 1288 CTTCCTGGCTCCATATCATATGAGGAGGCTTCTGGGTGCTTGAAGGGCAGCTTGT 1347  
Db 1103 CTTCCTGGCTCCATATCATATG----- 1123  
QY 1348 CTGCTACTGTGTGTGTGACATATATCCCAACAGAGGACATGAGAGGCTTAGATGATCC 1407  
Db 1124 ----- 1123  
QY 1408 CCAGGACATGAGAGGCTTAGATGATGATCCCATGATGAGAGGACATGATGTTGG 1467  
Db 1124 ----- 1123  
QY 1468 TGCCTTCTCCAGAGCGGCGTGTGTGCTGACCTGAAGCTCAGAGCCCAAGCCAC 1527  
Db 1124 -----CAGAGCGGCGGCTGTGTGTGCTGACCTGAAGCTCAGAGCCCAAGCCAC 1172  
QY 1528 GTGAGACAGCTTCTTGGCCCCCATGAGAGGATGAGCGAGCTTGTGTATGAGAAAGTC 1587  
Db 1173 GTGAGACAGCTTCTTGGCCCCCATGAGAGGATGAGCGAGCTTGTGTATGAGAAAGTC 1232  
QY 1588 CCTTGGCCCTTAATCTATGTTCCGCAATGATGTCGCGCTTGTACCTCATCATGATGTCGCA 1647  
Db 1233 CCTTGGCCCTTAATCTATGTTCCGCAATGATGTCGCGCTTGTACCTCATCATGATGTCGCA 1292  
QY 1648 TTTGTACTGCTACATGATGTGAGAGAACTGAATGAGCCAGGCCCTTCCGCAACCCCA 1707  
Db 1293 TTTGTACTGCTACATGATGTGAGAGAACTGAATGAGCCAGGCCCTTCCGCAACCCCA 1352  
QY 1708 GAATGTGACAGCTTCTTGGCGTGCACCTACGAGAGGCGCTTGTGACATGATGCTGGCT 1767  
Db 1353 GAATGTGACAGCTTCTTGGCGTGCACCTACGAGAGGCGCTTGTGACATGATGCTGGCT 1412  
QY 1768 GGGCCCACTCCAGTTCCTCATATAAAGACAGATGCTTCTGCTTCTGCTCTGACGAGG 1827  
Db 1413 GGGCCCACTCCAGTTCCTCATATAAAGACAGATGCTTCTGCTTCTGCTCTGACGAGG 1472  
QY 1828 GCCTTCGACATGAGTCTGGCCCTGGCCCACTCCCACTTCTCATATAAAGACAGAT 1887  
Db 1473 GCCTTCGACATGAGTCTGGCCCTGGCCCACTCCCACTTCTCATATAAAGACAGAT 1532  
QY 1888 TGGCTTCCTCACTTG 1901  
Db 1533 TGGCTTCCTCACTTG 1546

Db 207456 AT 207455

RESULT 10  
BC005395  
LOCUS  
DEFINITION Homo sapiens cDNA clone MGC:12533 IMAGE:4071616, complete cds.  
ACCESSION BC005395  
VERSION  
KEYWORDS  
SOURCE MGC.  
ORGANISM Homo sapiens (human)  
REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, J., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalski, U., Smailus, D.E., Schnerk, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 1621)  
Strausberg, R.  
Direct Submission  
Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLU)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdpaxil.stanford.edu](mailto:mdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.  
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMLU at: <http://image.llnl.gov>  
Series: IRAL Plate: 16 Row: h Column: 22.  
Location/Qualifiers  
1..1621  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:12533 IMAGE:4071616"  
/tissue\_type="Liver"  
/clone\_1ib="NIH\_MGC\_76"  
/lab\_host="DH10B"  
/note="Vector: PDNR-LIB"  
62..826

BASE COUNT 370 a 438 c 441 g 372 t  
ORIGIN  
Query Match 11.9%; Score 444.4; DB 9; Length 1621;  
Best Local Similarity 73.6%; Pred. No. 2,5e-128;  
Matches 721; Conservative 0; Mismatches 6; Indels 253; Gaps 2;  
/codon\_start=1  
/product="Unknown (protein for MGC:12533)"  
/protein\_id="AA05395.1"  
/db\_xref="GI:13529281"  
/translators="MARVELAPVALGMSLSLATPPLPTSGANNVABGETKDP  
DYTERCSDEWSEFATLDDNGTMLPFKGEFVMSHMDRELISERKKNFSPVDAAR  
OGHNSVFLIKGDVWVYPPBKKEKGPXKLODEFPPIPSFLDVAVECHRECAEGVL  
FPGDDEMFWDATGTMKERSMPAVCNCSALRWLGRVYCFQGNQFLRFPVGEVVP  
RYPRVDYDFMPGPRGHGRNGTGNISCAVAH"  
928 TCACAGGAGCCACTACTAGCGCTCTGACACACAGCCGGAGTGGCATAGCTGGCCC 987  
871 TTCACTGGGAGCCACTACTAGCGCTCTGACACACAGCCGGAGTGGCATAGCTGGCCC 930  
988 ATTGCTCATAGGCGCCCAAGGCTCTTACAGAGTGGAGTGGCTTTCTCTGGAGAA 1047  
931 ATTGCTCATAGGCGCCCAAGGCTCTTACAGAGTGGAGTGGCTTTCTCTGGAGAA 990  
1048 AAATCTATCTGCTCAGAGTGTATTGGGGAGAGGCTTAGGTAGAGACTGGACAAAG 1107  
991 AAATCTATCTGCTCAGAGTGTATTGGGGAGAGGCTTAGGTAGAGACTGGACAAAG 1003  
1108 CATATCAACTGTATTTATTATCAATCTTTGCTCCAGGACCCAGATATATGCTT 1167  
1004 TCCAGGACACCCAGATATATGCTT 1027  
1168 TCCAGCAAGAGGAGGCTTACCTAGTAAAGGCTTATCCGAGCGGCTGGAGAAAG 1227  
1028 TCTGTCAAGAGGAGGCTTACCTAGTAAAGGCTTATCCGAGCGGCTGGAGAAAG 1087  
1228 TCCGAGACCCCTCATAGGATTTATCTGAGTGGAGCGGCTTTATCTGCGCTGGGT 1287  
1088 TCCGAGACCCCTCATAGGATTTATCTGAGTGGAGCGGCTTTATCTGCGCTGGGT 1147  
1288 CTTCGCGCTCATATCATAGGAGGAGGCGCTTCTGGTGCTTAAAGGCAAGCTTGT 1347  
1148 CTTCGCGCTCATATCATAGGAGGAGGCGCTTCTGGTGCTTAAAGGCAAGCTTGT 1168  
1348 CTCTACCTGTCTGTGTGGCATAGATCCCAACAGGAGCATAGAAAGGCTTAGATCC 1407  
1169 1168  
1408 CCAGGACATGAGAGGCTTAGATCCCAACAGGAGCATAGAAAGGCTTAGATCC 1467  
1169 1168  
1468 TGCCTTCTCCAGAGCGGCGCTGTGTGGCTGACCTGAACTCAGAGCCCAAGCCAC 1527  
1169 -----CAGACGCGGCGCTGTGTGGCTGACCTGAACTCAGAGCCCAAGCCAC 1217  
1528 GTGAGACAGAGCTTCTTGGGCCCCCATAGAAAGTGAAGCGAGCTTGTGTATGAAAAAGTC 1587  
1218 GTGAGACAGAGCTTCTTGGGCCCCCATAGAAAGTGAAGCGAGCTTGTGTATGAAAAAGTC 1277  
1588 CTTGGCCCTTAATCATAGTTCGCGCAATGATCCCGGCTTGTACCTATCATATGATCCCA 1647  
1278 CTTGGCCCTTAATCATAGTTCGCGCAATGATCCCGGCTTGTACCTATCATATGATCCCA 1337  
1648 TTTGTACTGTACATGATGTGAGAAATGGAATGGAAGCAAGCCCTTCCGCAACCCA 1707  
1338 TTTGTACTGTACATGATGTGAGAAATGGAATGGAAGCAAGCCCTTCCGCAACCCA 1397  
1708 GAATGTACAGAGTCTCTGGGCTGACCTACATGAGAGGCGCTTGTACATGAGTGGCCT 1767  
1398 GAATGTACAGAGTCTCTGGGCTGACCTACATGAGAGGCGCTTGTACATGAGTGGCCT 1457  
1768 GGGCCACCTCTAGTTCCTCATTAATAAGACAGATTGCTTCTGCTTCTACATGAGG 1827





QY 1038 CTGGAGAGAAAACTATCTGGCCAGGCTGATTTGGGGAGAGGCTTGAGTTAGCA 1097  
 DB 173736 CTGGAGTTGAGAAAGTCTATCTGATTCAGGTTTGTATTTGGAGAAAGGCTTAAGATGGAAA 173795  
 QY 1098 CTGGAGCAGACATATCCACTCTGTATTTATTAACATCTTTGTCTCCAGGACCCAG 1157  
 DB 173796 TAGGAGCATGGCTAT--GGGTCTATGTATTAATCTCTCTTTGTCTCCAGGACACTCA 173853  
 QY 1158 GTTATGTCTTCTGACAAAGGAGTATACCTTAAGCGGTTATCCGAAGCGGCTG 1217  
 DB 173854 GTTATGTCTTCTGACAAAGGAGTATACCTTAAGCGGTTATCCGAAGCGGCTG 173913  
 QY 1218 GAGAGAGAGTGGGAGACCCCTCATGGGATATCTGTGAGCTGTGAGATGGCGCTTATC 1277  
 DB 173914 GAGAGAGATTTGGAGACCTTCCCGGATCAGCTTGTATATCATAGATGAGGCTTTTCC 173973  
 QY 1278 TGGCTTGGCTCTTCTGGCTCATATCATGAGGAGGAGGCTTGTGGTGGCTTAGAG 1337  
 DB 173974 TGGCTTGGCTCTTCTGGCTCATATCATGAGGAGGAGGCTTGTGGTGGCTTAGAG-GG 174032  
 QY 1338 GCAAGCTTTTCTGCTACTGTCTGTGAGATATCCCAACAGGAGGAGGCTTAG 1397  
 DB 174033 GTGTCTTGTCTTCTTAATCTTCAACAGCATAGAT----- 174066  
 QY 1398 GTGAGATCCCGAGGAGTGAAGAGGCTTAGGCTAGGATCCCATGACATGAGGAGCATG 1457  
 DB 174067 -----GATCAGAGAGGAGTGAAGAGGCTTAGGCTAGGATCCCATGACATGAGGAGCATG 174121  
 QY 1458 CTATGTTTGTGCTTCTTCCCAAGAGCGGCGCTGTGTGCTGAGACTGAGAGTCAAGAG 1517  
 DB 174122 TTAAAGTTTGTGCTTCTTCCCAAGAGCGGCGCTGTGTGCTGAGACTGAGAGTCAAGAG 174181  
 QY 1518 CCCAAGCCAGTGAAGAGAGCTTCTTGGCCCATGAGAAAGTGAAGACGAGCCTTGTGA 1577  
 DB 174182 CCCAAGCCAGTGAAGAGAGCTTCTTGGCCCATGAGAAAGTGAAGAGCCTTGTGT 174241  
 QY 1578 TGGAAAGTCCCTGGCCCTACTCATGTTCCGCAATGGTCCCGGCTTACTCATCC 1637  
 DB 174242 TGGAAAGTCCCTGGCCCTACTCATGTTCCGCAATGGTCCCGGCTTACTCATCC 174301  
 QY 1638 ATGTGCTCAATTTGATCTGCTAGACATGATGAGAACTGAATGAGCCCAAGCCCTTC 1697  
 DB 174302 ATGTGCTCAATTTGATCTGCTAGACATGATGAGAACTGAATGAGCCCAAGCCCTTC 174361  
 QY 1698 CGAAGCCCAAGATGAGACAGTCTCTTGGGCTGCACTCACTGAGGAGCC--TTCTGAC 1754  
 DB 174362 CTCAGGCCCAAGAAAGTGAACAGCATCTTGGCTGCACTCACTCAATAAAAAGCCCTGATGAG 174421  
 QY 1755 ATGAGTCTGGCCCTGGCCCACTCTGATGTTCCCTAATTAAGACAGATTGCTTCCG 1814  
 DB 174422 TTAGCCAGCCAGCCCACTCTGATTTCCATTTAATAAACAAGATGTTTCTTCA 174481  
 QY 1815 TT 1816  
 DB 174482 AT 174483

RESULT 9  
 AC131626 275982 bp DNA linear HTG 10-MAY-2003  
 AC131626/c  
 LOCUS Rattus norvegicus clone CH230-163P16, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 15 unmerged pieces.  
 AC131626  
 ACCESSION AC131626 GI:30520421  
 VERSION AC131626.5 HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
 1 (bases 1 to 275982)  
 REFERENCE

REFERENCE  
 JOURNAL  
 TITLE  
 AUTHORS  
 JOURNAL  
 COMMENT  
 REFERENCE  
 JOURNAL  
 TITLE  
 AUTHORS  
 JOURNAL  
 COMMENT

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anylebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bunay, C., Burch, P., Butrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyler, M., Crete, A., D'Souza, L.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregregis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,  
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenschew, L., Louieged, H., Lozano, R.J., Lu, X., Ma, D.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, B.,  
 Milobav, J., A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norrie, S.,  
 Nwackelme, O., Okunolu, G., Olampunsgoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, F., Polidexter, A., Popovic, D., Prims, B., Pu, L., L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, S., Reich, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, R., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaj, D.,  
 Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Soza, P.,  
 Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villaseana, D., Waldon, L., Walker, B., Wang, J.,  
 Wang, O., Wang, S., Warren, J., Wleczek, R., Wood, H., Worley, K.,  
 Williams, G., Willson, R., Wleczek, R., Wood, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, Y., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G. and Gibbs, R.A.  
 Direct Submission  
 Title  
 Unpublished  
 2 (bases 1 to 275982)  
 Reference  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (25-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 275982)  
 Reference  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:25097807.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome

Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bismuelo, K., Blair, J., Blankenburg, K., Blych, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Creg, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Herliand, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howell, S., Huily, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbuewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munkidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelam, O., Okunolu, G., Olarunmugoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puato, M., Quiriz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Shetty, J., Shvartbeyan, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, Y., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 238748)  
Worley, K.C.

Direct Submission  
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 238748)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23265319. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GXYV  
Center clone name: CH230-25H1  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 220383 bases at least Q40  
Consensus quality: 22613 bases at least Q30  
Consensus quality: 226021 bases at least Q20  
Estimated insert size: 223050; sum-of-estimates  
Quality coverage: 7x in Q20 bases; sum-of-estimates  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 238748: contig of 238748 bp in length.  
Location/Qualifiers  
1. 238748  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-25H1"  
1904. 2550  
/note="clone\_boundary  
clone\_end:77  
site:  
end\_sequence: BH303560"  
misc\_feature  
43040. 44214  
/note="wgs config"  
complement(124965. 236969)  
/note="clone\_boundary  
clone\_end:5p6  
site:  
end\_sequence: BH303561"  
142512. 143632  
/note="wgs and extension  
clone\_end:5p6"  
misc\_feature  
64699 a 51467 c 52376 g 59689 c 10517 others  
ORIGIN  
Query Match 13.9%; Score 520; DB 2; Length 238748;  
Best Local Similarity 73.3%; Pred. No. 7.6e-152;  
Matches 749; Conservative 0; Mismatches 235; Indels 38; Gaps 5;  
-----  
QY 798 CTTTCACCTGGCCTTTCATCTTGGCCTTGATGATCTTCTCTCATGACTCA 857  
DB 173497 CTTTTCACCTGACCTTCTTCTTGGCCCTCGATACATCATCTTATGACTCA 173556  
QY 858 TTTTCGATTCATCATGACCTTCTCTGCGCTTCTGCGGCTTCTGCGAGCCCTTGAAGCAA 917  
DB 173557 AACCTACCTCCATTTCTGCTT-TTCACATGAAAGTCTTACACCACTGGACATTA 173615  
QY 918 CCTATGATTTTCACAGGAGCCACTACTAGGGCTGTGACACACAGCCGGATGCTGGCA 977  
DB 173616 CCTTGATATCCACAGGCTCCCATCTAGGGCTGTGACCTCAAGCCGATGAGTGGCA 173675  
QY 978 TAGCTGGCCCATTTGCTCATCACTGAGCCCAAGGCTCTTTCAGCAGTGAATGCTGCTTTTC 1037  
DB 173676 TAGCTGGCCCATTTGCTCATCACTGAGCCCAAGGCTCTTTCAGCAGTGAATGCTGCTTTTC 173735

QY 1756 TGAAGTCGAGCTGGCCCACTCTAGTTCCTCATATAAAGACAGATTGCTTGGCT 1815  
 DB 131078 TGAATATGGCCAGCCCACTCTCATATTTCATCTAATAAGCAATGTTTCTTCA 131137  
 QY 1816 T 1816  
 DB 131138 T 131138

RESULT 7  
 MMEPEX8 1467 bp DNA linear ROD 29-JUL-2002  
 LOCUS M.musculus gene for hemopexin (exons 8-10).  
 DEFINITION X56830  
 ACCESSION X56830.1 GI:22022647  
 VERSION haem binding; hemopexin; plasma glycoprotein.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 372 to 1467)  
 AUTHORS Alam, J. and Smith, A.  
 TITLE Nucleotide sequence of the mouse haemopexin gene  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1467)  
 AUTHORS Alam, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-1990) Alam J., Alton Ochsmar Medical Foundation,  
 Department of Molecular Genetics, 1516 Jefferson Highway, New  
 Orleans, LA 70121, USA  
 FEATURES  
 source 1..1467  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>1467  
 /number=9  
 intron join(461..591,675..837,975..1234,461..591,675..837,  
 975..>1234)  
 /usedin=X56829:mRNA\_hmpex  
 exon 461..591  
 /number=8  
 intron 592..674  
 /number=7  
 exon 675..837  
 /number=9  
 intron 838..974  
 /number=8  
 exon 975..>1234  
 /number=10  
 BASE COUNT 359 a 364 c 349 g 395 t  
 ORIGIN

Query Match 14.9% Score 557.8; DB 10; Length 1467;  
 Best Local Similarity 73.8%; Pred. No. 3e-164; Indels 40; Gaps 2;  
 Matches 754; Conservative 0; Mismatches 227;

QY 796 TCCTTCTCACTGGCTCTTCATCTTGAGTCATCTCTCCCTCATGACT 855  
 DB 322 TTCACACTATTCACCTCTTCCTGAGGCTCTGATGCATTCCTCTCATGACT 381  
 QY 856 CATTTCGATTCATCACTAGCCTCTTCTGCTGGCTTTCGCAAGGCCCTAGAC 915  
 DB 382 CAACCTGATCATCTGTCCTCTTCACTAGAAAGCTTTACCAACAGCACTGAC 441  
 QY 916 AACCTATGATTTCCACAGAGAACCACTAGCGCTGAGACACAGCCGGATGCTGG 975  
 DB 442 AACCTTATCACTCCACAGAGCTTCACACTGAGCTGCTGACCTCCAGCCGTATGATG 501  
 QY 976 CATAGCTGACCATGCTCATAGTGGCCCAAGGCTCTTCAGCAAGTGAATGCTGCTTT 1035

DB 502 CATAGCTGACCATTCCTCATCACTGGCCCTCAAGGTCCTTCAACAGTAGATGCTGCTTT 561  
 QY 1036 TCCTGGAGAAATACTATCTATGTCAGAGTGTATTTGGGGAGAGGCTTAGAGTAGA 1095  
 DB 562 TCCTGGAGATTAAGTCTATCTGATCGATGATGATGAGGAGAAAGCTTAAGATGGA 621  
 QY 1096 GACTGGACAAGCATATCCAACTGTATTTATTAACATCTTTTGTCTTCAGAGGACCC 1155  
 DB 622 AACAGGACAAGGCTATACAGGTCTATGTGATTAATCTCTCTTGTTCCTCCAGGCACTC 681  
 QY 1156 AGGTATATGTCTTCTGACAAAGGAGGCTATACCTAGTAAGCGGTATCCGAGCGGC 1215  
 DB 682 AAGTATATGTCTTCTGACAAAGGAGGCTATACCTAGTAAGTGTATCCAAAGCGGC 741  
 QY 1216 TGGAGAAAGAAAGTCGGGACCCCTCAGGATTAATCTGAGACTCTGAGATGGGCTTTA 1275  
 DB 742 TGGAGAAAGAACTGGGAGCCCTCTGGGATCAGCTTGAACATAGATAGACCTTTT 801  
 QY 1276 TCTGACCTGGGCTTCTCGGCTCCATATCATGAGGAGGAGGCTTCTGGTGTAG 1335  
 DB 802 CCTGACCTGCTTCTTCAGAGCTCTATGCTCATCAGTGAAGTCTTTGGGTACTAAG 861  
 QY 1336 GGGCAGCTTGTCTGCTACTGTCTGTGCAATGATCCGACCAAGGCGATGAGAGGCTT 1395  
 DB 862 GATGCTTGTCTTTAACTTTTCAAGGATA----- 892  
 QY 1396 AGGTACAGATCCCAAGGAGATAGAAAGGCTTAGTCAAGATCCCAATGACAGCA 1455  
 DB 893 ---GATGACCAAGAGGAGCATTAAGAACTTAATCAAGTCTGCAATGATAGAACTA 948  
 QY 1456 TGTATGTTGGTGCCTTCTCCCAAGAGCGGCGCTGTGTGGCTGAGCTTAAGTCAAG 1515  
 DB 948 TATTAAGTTTGTGTCTTCTTCCAGAGAGGCGGCTTGTGTGCTGACTGAAGTCAAG 1008  
 QY 1516 AGCCCAAGCAGCTGAGACAGAGCTTCTTGGCCCAATGAAAGTACGAGAGCTTGTG 1575  
 DB 1009 AGCCCAAGCAGATGAGACAGAGTTCCTGGCCCAATGAAAGTATGAGGCGCTGTG 1068  
 QY 1576 TATGAAAGTCCCTTGGCCCTTAATCAATGTTCCGCAATGTCTCCGCTTGTACTCAT 1635  
 DB 1069 TTTGACAAAGTCTTGGCCCAACAGATTTTCCAAATGTTCCAGCTTGTACTTAT 1128  
 QY 1636 CCATGCTCCCAATTTGATCTGCTACAGTATGAGAAATGAATGACAGCAAGGCTT 1695  
 DB 1129 CCATGCGCCCAATTTGATCTGCTATGACAGTATGACAAATGATGACAGCAAGGCTT 1188  
 QY 1696 TCCGCAACCCCAAGATGATGACAGTCTCTGAGCTGACCTAGAGGGGCTTGTACA 1755  
 DB 1189 GCCTCAGCTCAGAAAGTGAACAGATCTTGGCTGACAGTAAAGAGC-----CC 1241  
 QY 1756 TGAAGTCGAGCTGGGCCCCACCTCTAGTTCCTCATATAAAGACAGATGCTTTCGCT 1815  
 DB 1242 TGAATATGGCCAGCCCACTCTTCATCTTAATAAGCAAGATGTTTCTTCA 1301  
 QY 1816 T 1816  
 DB 1302 T 1302

RESULT 8  
 AC120746 238748 bp DNA linear HTG 19-NOV-2002  
 LOCUS Rattus norvegicus clone CH230-25H1, WORKING DRAFT SEQUENCE.  
 DEFINITION AC120746  
 AC120746.4 GI:25086001  
 VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS \_FULLTOP.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 238748)  
 AUTHORS Muzny, D., Marie, M., Metzger, M., Lee, S., Abiraman, S., Adams, C., Alder, J.,

```
* 5211 7669: contig of 2459 bp in length
* 7670 7669: gap of unknown length
* 7770 13621: contig of 5852 bp in length
* 13622 13721: gap of unknown length
* 13722 19656: contig of 5935 bp in length
* 19657 19756: gap of unknown length
* 19757 31061: contig of 11305 bp in length
* 31062 31161: gap of unknown length
* 31162 40106: contig of 8945 bp in length
* 40107 40206: gap of unknown length
* 40207 50613: contig of 10407 bp in length
* 50614 50713: gap of unknown length
* 50714 61952: contig of 11239 bp in length
* 61953 62052: gap of unknown length
* 62053 74978: contig of 12926 bp in length
* 74979 75078: gap of unknown length
* 75079 92714: contig of 17636 bp in length
* 92715 92814: gap of unknown length
* 92815 124084: contig of 31270 bp in length
* 124085 151377: contig of 27193 bp in length
* 151378 151477: gap of unknown length
* 151478 178112: contig of 26635 bp in length
* 178113 178212: gap of unknown length
* 178213 208382: contig of 30170 bp in length
* 208383 208482: gap of unknown length
* 208483 249448: contig of 40966 bp in length
* 249449 249548: gap of unknown length
* 249549 249622: contig of 74 bp in length
* 249623 249723: gap of unknown length
* 249723 249769: contig of 47 bp in length.
```

## FEATURES

source

```
1. 249769
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="X"
  /clone="RP23-42113"
  1. 1085
  /note="assembly_name:Contig33"
  1186. 3097
  /note="assembly_name:Contig84"
  3198. 5110
  /note="assembly_name:Contig85"
  5211. 7669
  /note="assembly_name:Contig86"
  7770. 13621
  /note="assembly_name:Contig87"
  13722. 19656
  /note="assembly_name:Contig88"
  19757. 31061
  /note="assembly_name:Contig89"
  31162. 40106
  /note="assembly_name:Contig90"
  40207. 50613
  /note="assembly_name:Contig91"
  50714. 61952
  /note="assembly_name:Contig92"
  62053. 74978
  /note="assembly_name:Contig93"
  75079. 92714
  /note="assembly_name:Contig94"
  92815. 124084
  /note="assembly_name:Contig95"
  124185. 151377
  /note="assembly_name:Contig96"
  151478. 178112
  /note="assembly_name:Contig97"
  178213. 208382
  /note="assembly_name:Contig98"
  208483. 249448
  /note="assembly_name:Contig99"
  249549. 249622
```

```
misc_feature /note="assembly_name:Contig24"
249723. 249769
/note="assembly_name:Contig95"
BASIC COUNT 67104 a 56647 c 55906 g 68301 t 1011 others
ORIGIN
Query Match 15.2%; Score 567.4; DB 2; Length 249769;
Best Local Similarity 75.0%; Pred. No. 7.5e-167;
Matches 766; Conservative 0; Mismatches 216; Indels 39; Gaps 3;
QY 796 TCCTTTGACGCGCTCTTCATCTGGCTCTGGATGATCTCTCCCTCATGACT 855
DB 130157 TTCACCTGATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 130216
QY 856 CATTTGATTCATCATCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 915
DB 130217 CAACCTGATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 130276
QY 916 AACCTATGATTCATCAAGGACCCCACTGCGCTCTGACACACCGCGGATGCTG 975
DB 130277 AACCTATGATTCATCAAGGACCCCACTGCGCTCTGACACACCGCGGATG 130336
QY 976 CATAGTGGCCCATTTGCTCATGATGCGCCCAAGGCTCTTCAAGAGTGAATGCT 1035
DB 130337 CATATCTGGCCCATTTGCTCATGATGCGCCCAAGGCTCTTCAAGAGTGA 130396
QY 1036 TCCTGGGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1095
DB 130397 TCCTGGGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 130456
QY 1096 GACTGGGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1155
DB 130457 AACGGGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130516
QY 1156 AGGTATATGCTCTTCTGACAAAGGAGGCTATACCTAGTAAGCGGTATCCGA 1215
DB 130517 AGGTATATGCTCTTCTGACAAAGGAGGCTATACCTAGTAAGCGGTATCC 130576
QY 1216 TGGAGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1275
DB 130577 TGGAGAGAGAAACCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 130636
QY 1276 TCTGCCCTGGCTCTTCTCGGCTCCATATGATGAGGAGGCTCTTCTGCTGCT 1335
DB 130637 CCGCCCTGGCTCTTCTCGGCTCCATATGATGAGGAGGCTCTTCTGCTGCT 130695
QY 1336 GGGCAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1395
DB 130696 GGGCAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130755
QY 1396 AGGTACGATTCCTGAGGAGGATGAGAAAGGCTAGTCCGATCCGATGAGAGCC 1455
DB 130756 AATTCAGGCTG-----CATGATGAGAAACTG 130784
QY 1456 TGTATGTTGGTGGCTTCTCCGAGAGCGGCTGTGCTGCTGCTGCTGCTGCTGCT 1515
DB 130785 TATTAAGTTGTGTTCTTCTCCGAGAGCGGCTGTGCTGCTGCTGCTGCTGCT 130844
QY 1516 AGCCCAAGCAGTGGACAGAGCTTCTTGGCCCAAGAGAGAGTGAAGAGCTTGTG 1575
DB 130845 AGCCCAAGCAGTGGACAGAGCTTCTTGGCCCAAGAGAGAGTGAAGAGCTT 130904
QY 1576 TATGAGAAAGTCCCTTGGCCCAAGTTCCTGCGCAATGATCCGCGCTGATCTCT 1635
DB 130905 TTTGACAGATCTCTTGGCCCAAGTTCCTGCGCAATGATCCGCGCTGATCTCT 130964
QY 1636 CCATGCTCCCAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1695
DB 130965 CCATGCTCCCAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 131024
QY 1696 TCCGCAACCCGAAATGATGACAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1755
DB 131025 GCCTCAGCTCAGAAAGTGAACAGATCTCTGCTGCTGCTGCTGCTGCTGCTG 131077
```



Db 22348 CATGCTTGTTGTCATGATGATGAGCATGAGTAATGAATGAGTTGAGTGGGAAAC 22407  
QY 2400 AGAATATACCCAGCTGCTGTCAGAGCAGCTGAGAGAGAGA--GAAAGATGAATG 2457  
Db 22408 AGAATATACCCAGCTGCTGTCAGAGCAGCTGAGAGAGAGA--GAAAGATGAATG 22467  
QY 2458 GCAACATGAGACATTTGAGAGACACAGTCCCTGAGAGCTTTGCTGAGAGCTCAGAGC 2517  
Db 22468 GCAACATGAGACATTTGAGAGACACAGTCCCTGAGAGCTTTGCTGAGAGCTCAGAGC 22527  
QY 2518 TTT 2520  
Db 22528 TCT 22530

RESULT 4  
HUMHXA07 629 bp DNA linear PRI 31-DEC-1994  
LOCUS Human hemopexin gene, exon 10.  
DEFINITION M36803  
ACCESSION M36803.1 GI:184495  
VERSION hemopexin.  
KEYWORDS 7 of 7  
SEGMENT Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Altruda, F., Pol, V., Restagno, G. and Silengo, L.  
TITLES Structure of the human hemopexin gene and evidence for intron-mediated evolution  
JOURNAL J. Mol. Evol. 27 (2), 102-108 (1988)  
MEDLINE 88316972  
PUBMED 2842511  
COMMENT Original source text: Homo sapiens DNA.  
FEATURES  
source location/Qualifiers  
1. 629  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="1p15.5-p15.4"  
/join("M36796.1:178..731,M36798.1:1..457,M36799.1:1..278,M36800.1:1..404,M36801.1:1..278,M36802.1:1..577,1..468)  
/gene="HPX"  
/join("M36796.1:206..288,M36796.1:439..497,M36796.1:618..689,M36798.1:29..150,M36799.1:109..262,M36800.1:81..293,M36801.1:106..237,M36802.1:118..248,M36802.1:332..494,50..309)  
/gene="HPX"  
/note="precursor"  
/codon\_start=1  
/product="hemopexin"  
/protein\_id="AA58678.1"  
/db\_xref="GI:184497"  
/db\_xref="GDB:G00-120-054"  
/translation="MARVLGAPVALGMSLCSLATATPLPPTSAHGNVAREGTRDPDYERCSIDMSPTATLDNGTMLPFGGEFVMSKMDRELISERRKPSPPDAAPRQGNISVPLIGDKVWYTPPEKSKYFKLQDEPFPPIPSLDAVNECHRESCAEGVLFQGDREWFMDLAVGTWKEKSPAVGSCSLALMGLRYCFQNGFLRDPVAGEVPRYPRVDYDFMPCPGHGHIRNGTGHNSHTHGPPEYKCPHLVLSALISDNHATYAFSGTHYRLDTSRDGMSHWPISAHQWPGSPAVAAFSWEKLYLVGTQVYVFLTKGGYTLVSGYPRKLEKEVGTPIGIIIDSVDAATICGSSRLHIMAGRLMMLDLKGAOATWELPMPHEKVDALCMKESLGNCSANPGIYLHGPVLYCSVDEKLNAAKALPQPOVNTSLGCTH"  
mat\_peptide  
join("M36796.1:275..288,M36796.1:439..497,M36796.1:618..689,M36798.1:29..150,M36799.1:109..262,M36800.1:81..293,M36801.1:106..237,M36802.1:118..248,M36802.1:332..494,50..306)  
/gene="HPX"  
/product="hemopexin"  
/note="G00-120-054"  
/gene="HPX"  
intron  
<1..49  
/gene="HPX"

/note="G00-120-054"  
/number=9  
50..468  
/gene="HPX"  
/note="G00-120-054"  
/number=10  
442..447  
/gene="HPX"  
/note="G00-120-054"  
BASE COUNT 142 a 172 c 159 g 156 t  
ORIGIN About 1.5 kb after segment 6.  
Query Match 16.5%; Score 618; DB 9; Length 629;  
Best Local Similarity 99.8%; Pred. No. 2,2e-183;  
Matches 629; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1433 GATATCCCATGACATGAGAACCATGCTATGTTGCTTCTCCAGAGACGGCGCTG 1492  
Db 1 GATATCCCATGACATGAGAACCATGCTATGTTGCTTCTCCAGAGACGGCGCTG 60  
QY 1493 TGTGCTGACCTGTAAGTCAGAGAGCCCAAGCCAGGTGAGACAGCTTCTGCGCCAT 1552  
Db 61 TGTGCTGACCTGTAAGTCAGAGAGCCCAAGCCAGGTGAGACAGCTTCTGCGCCAT 120  
QY 1553 GAGAAAGTAGACGAGAGCTTGTGTATGAAAGTCCCTTGCCCTTAATCATGTTCCGCC 1612  
Db 121 GAGAAAGTAGACGAGAGCTTGTGTATGAAAGTCCCTTGCCCTTAATCATGTTCCGCC 180  
QY 1613 AATGTCCTGCTGTTGACTCATTCATGATGTTCCCAATTGTTGTTACTGTTACATGATGTTGAG 1672  
Db 181 AATGTCCTGCTGTTGACTCATTCATGATGTTCCCAATTGTTGTTACTGTTACATGATGTTGAG 240  
QY 1673 AAATGTAATGACGAGAGCCCTTCCGCAACCCAGAAATGTAACGATGTTCCCTGCGGCTGC 1732  
Db 241 AAATGTAATGACGAGAGCCCTTCCGCAACCCAGAAATGTAACGATGTTCCCTGCGGCTGC 300  
QY 1733 ACTCATGAGAGGCTTCTGACATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1792  
Db 301 ACTCATGAGAGGCTTCTGACATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 359  
QY 1793 TAAAGACATGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1852  
Db 360 TAAAGACATGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419  
QY 1853 CCCACCTCCCAAGTTCATTAATAAGACGATGCTTCTTCACTGTAATCAAGGAGAC 1912  
Db 420 CCCACCTCCCAAGTTCATTAATAAGACGATGCTTCTTCACTGTAATCAAGGAGAC 479  
QY 1913 CTTGCTGTAACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1972  
Db 480 CTTGCTGTAACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 539  
QY 1973 TGTGAGCTCAACAGAGCTGTGAGAACAGAGGAGGAGACATTAAGGAGCAACCTAT 2032  
Db 540 TGTGAGCTCAACAGAGCTGTGAGAACAGAGGAGGAGACATTAAGGAGCAACCTAT 599  
QY 2033 CTTGCTGAGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2062  
Db 600 CTTGCTGAGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 629

RESULT 5  
HUMHXA06 577 bp DNA linear PRI 31-DEC-1994  
LOCUS Human hemopexin gene, exons 8 and 9.  
DEFINITION M36802  
ACCESSION M36802.1 GI:184494  
VERSION hemopexin.  
KEYWORDS 6 of 7  
SEGMENT Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Primates; Catarrhini; Homnidae; Homo.



Qy	18	AAAAAGCTAAGAGAAATACCTGGGAGGTGACGGGTGCACGAAGCCAAAGTCCAAAAAA	240
Db	20188	AAAAAGATAGAGAAATACCTGGGAGGTGACGGGTGCACGAAGCCAAAGTCCAAAAAA	20247
Qy	241	GACATTTAAAGGAGAAAGAGTAGTAGCGAGTCCAGTCCGTCTGAGAGGTAGGGTCAGAT	300
Db	20248	GACATTTAAAGGAGAAAGAGTAGTAGCGAGTCCAGTCCGTCTGAGAGGTAGGGTCAGAT	20307
Qy	301	GAGAACGAGAAATTGACCATGAGATTTCCGAATTTGGAGAAATCTAGCAACTGGATTAAG	360
Db	20308	GAGAACGAGAAATTGACCATGAGATTTCCGAATTTGGAGAAATCTAGCAACTGGATTAAG	20367
Qy	361	AACAATTTCAATGTTGAGGGGAAACAGAACTGTAAATTTGAAGAGAGATTTGAGGAAAAAGAC	420
Db	20368	AACAATTTCAATGTTGAGGGGAAACAGAACTGTAAATTTGAAGAGAGATTTGAGGAAAAAGAC	20427
Qy	421	AAATGGAGGCTAGATTAATTCCTTAATAAGTTGTGAAAAGAGAGAAAGAAAAACGGG	480
Db	20428	AAATGGAGGCTAGATTAATTCCTTAATAAGTTGTGAAAAGAGAGAAAGAAAAACGGG	20488
Qy	481	GTCGTAGCCCAAGCTACTCTCCCTCACTCTTCCACAACCTCATAGGAGAGACTGGAGAACAC	540
Db	20488	GTCGTAGCCCAAGCTACTCTCCCTCACTCTTCCACAACCTCATAGGAGAGACTGGAGAACAC	20547
Qy	541	AGCCAGAGTGAAGAACTTCAGTGAAGAGTGGTGGTCTTTTAATGTTTGGAGACATGTAT	600
Db	20548	AGCCAGAGTGAAGAACTTCAGTGAAGAGTGGTGGTCTTTTAATGTTTGGAGACATGTAT	20607
Qy	601	TTCAATTAATTAACCCGATCTCTGTACATGAGACACCTGAATTCCTTAAAGGAGTGGCCGC	660
Db	20608	TTCAATTAATTAACCCGATCTCTGTACATGAGACACCTGAATTCCTTAAAGGAGTGGCCGC	20667
Qy	661	CAACCCCATGATGTTGGCTTACCTGGAACCTTAGCCACTGTTTTCACAACCTTGCCTTTC	720
Db	20668	CAACCCCATGATGTTGGCTTACCTGGAACCTTAGCCACTGTTTTCACAACCTTGCCTTTC	20727
Qy	721	TTTTAGGCACTGGTGATTCAGATTTCAAGGCAAGGAGCAAGTGGCCAAATGGTGACAA	780
Db	20728	TTTTAGGCACTGGTGATTCAGATTTCAAGGCAAGGAGCAAGTGGCCAAATGGTGACAA	20787
Qy	781	GTCCTTGCTAATTTCTCTCTCTCACCTGGGCTCTTCCATCTTGGGCTCTGAGATGCAATCT	840
Db	20788	GTCCTTGCTAATTTCTCTCTCTCACCTGGGCTCTTCCATCTTGGGCTCTGAGATGCAATCT	20847
Qy	841	CTCCCTCTCAATGACTAATTTCTGCAATTCACACTAGGCTCTTCTGCTGGGCTTCTGCG	900
Db	20848	CTCCCTCTCAATGACTAATTTCTGCAATTCACACTAGGCTCTTCTGCTGGGCTTCTGCG	20907
Qy	901	CAGGGGCTTGAAGCAACTATGATTTCCACAGGAGCCCACTACTGGGCTCTGAGACC	960
Db	20908	CAGGGGCTTGAAGCAACTATGATTTCCACAGGAGCCCACTACTGGGCTCTGAGACC	20967
Qy	961	AGCCGGGATGGCTGAGCTAGCTGGCCATTTGCTCATCAAGTGGGCCAGGAGGCTCTTCAGCA	1020
Db	20968	AGCCGGGATGGCTGAGCTAGCTGGCCATTTGCTCATCAAGTGGGCCAGGAGGCTCTTCAGCA	21027
Qy	1021	GTEGATGCTGCTTTTCTGGGAGAAAACTCTATCTGGTCCAGGCTGTGTATTGGGGGA	1080
Db	21028	GTEGATGCTGCTTTTCTGGGAGAAAACTCTATCTGGTCCAGGCTGTGTATTGGGGGA	21087
Qy	1081	GAGGCTTAGAGTAGAGCTGGGACAGCAATATCCACTGTATTTATTAACATCCTTGG	1140
Db	21088	GAGGCTTAGAGTAGAGCTGGGACAGCAATATCCACTGTATTTATTAACATCCTTGG	21147
Qy	1141	TTCCTCAGGGGCACCAGATATATGTCTTCTGCAAAAGGAGGCTATACCCTTAGAGGG	1200
Db	21148	TTCCTCAGGGGCACCAGATATATGTCTTCTGCAAAAGGAGGCTATACCCTTAGAGGG	21207
Qy	1201	GTTATCCGAAGCGGCTGGAGAGAAAGTCGGGACCCCTCAGTGGGATTAATCTGGACTCTG	1260
Db	21208	GTTATCCGAAGCGGCTGGAGAGAAAGTCGGGACCCCTCAGTGGGATTAATCTGGACTCTG	21267
Qy	1261	TGAGTGGGCGCTTTATCTGCTTGGGCTTCTGAGCTCCATATCATGGCAGGTGAAGGGC	1320

Db	21268	TGATGCGGCGCTTTATCTGCGCTCGAGTCTTCTCGAGCTCAATATCATGGAAGGTGAGGGGC	2132
Qy	1321	TTCTGGGTGCTTAGAGGGCAGCTTGTTCTGCTACCTGTCTGTGGCATAGATCCCAACAG	1380
Db	21328	TTCTGGGTGCTTAGAGGGCAGCTTGTTCTGCTACCTGTCTGTGGCATAGATCCCAACAG	2138
Qy	1381	GGCATGAGAAAGGCTTAGGTCAAGATCCCCAGGGCATAGAAAGGCTTAGGTCAAGATCCCC	1440
Db	21388	GGCATGAGAAAGGCTTAGGTCAAGATCCCCAGGGCATAGAAAGGCTTAGGTCAAGATCCCC	2144
Qy	1441	ATGACATGGAAGCCATGCTATGTTTGGTGCCTTCCCGCAGAGCGGGGCGTGTGGTCT	1500
Db	21448	ATGACATGGAAGCCATGCTATGTTTGGTGCCTTCCCGCAGAGCGGGGCGTGTGGTCT	2150
Qy	1501	GGACCTGAAGTCAGAGAGCCCAAGCCAGTGGACAGAGCTTCTGTGGCCCATAGAGAGGT	1560
Db	21508	GGACCTGAAGTCAGAGAGCCCAAGCCAGTGGACAGAGCTTCTGTGGCCCATAGAGAGGT	2156
Qy	1561	AGACGAGAGCTTGTGTATGAGAAAATCCCTTGGCCCTTACTCATATGTTCCGCAATGCTCC	1620
Db	21568	AGACGAGAGCTTGTGTATGAGAAAATCCCTTGGCCCTTACTCATATGTTCCGCAATGCTCC	2162
Qy	1621	CGGCTGTGAACCTCATGATGCTGCCAATTTGTACTGTGTAACAGATGTGGAGAACTGAA	1680
Db	21628	CGGCTGTGAACCTCATGATGCTGCCAATTTGTACTGTGTAACAGATGTGGAGAACTGAA	2168
Qy	1681	TGCAGCCAAAGGCCCTTCCGCAACCCAGAAATGTGACCAAGTCTCCTGGGCTGCACTCACTG	1740
Db	21688	TGCAGCCAAAGGCCCTTCCGCAACCCAGAAATGTGACCAAGTCTCCTGGGCTGCACTCACTG	2174
Qy	1741	AGGGGCGCTTGTGACATAGTCTGGCTGCGCCCACTCTTAGTTCCTCATATTAAGACA	1800
Db	21748	AGGGGCGCTTGTGACATAGTCTGGCTGCGCCCACTCTTAGTTCCTCATATTAAGACA	2180
Qy	1801	GATTGCTTCTTGGCTTCTCATGAGGGGGCCCTCTGACATAGT-CTGGAGCTGGCCCCACC	1859
Db	21808	GATTGCTTCTTGGCTTCTCATGAGGGGGCCCTCTGACATAGTCTGGAGCTGGCCCCACC	2186
Qy	1860	TCCCCAGTTCCTCATATTAAGACAGATTGCTTCTTCACTTGAATCAAGGACCTTGGTC	1919
Db	21868	TCCCCAGTTCCTCATATTAAGACAGATTGCTTCTTCACTTGAATCAAGGACCTTGGTC	2192
Qy	1920	GTCGAAACAATCTTCTTCTTGTGAGTTGAAAAGTTAGACATTCCTCTTGGAGGTGTGAG	1979
Db	21928	GTCGAAACAATCTTCTTCTTGTGAGTTGAAAAGTTAGACATTCCTCTTGGAGGTGTGAG	2198
Qy	1980	CTCAAAACAAGGCTGTGAGAACAAAGGAGGGGAGCACTAAGGGGCAAACTATCTCTGCG	2039
Db	21988	CTCAAAACAAGGCTGTGAGAACAAAGGAGGGGAGCACTAAGGGGCAAACTATCTCTGCG	2204
Qy	2040	CAGATGATTTCTTAGAGTCAAGATCATAACTAGCTCTTGTGACACTATCTTCACATAGTGG	2099
Db	22048	CAGATGATTTCTTAGAGTCAAGATCATAACTAGCTCTTGTGACACTATCTTCACATAGTGG	2210
Qy	2100	GGGGAAAGAGAACCAAGTCTGGAAGAGAAACAGCTGATTTATACAGCAAGTAAGAGTG	2159
Db	22108	GGGGAAAGAGAACCAAGTCTGGAAGAGAAACAGCTGATTTATACAGCAAGTAAGAGTG	2216
Qy	2160	GAGCTTAGAGCTTGATTCGAACTTGTCTGTGATAGTGGCCAAACCCAGCCGCAAGGCATCAG	2219
Db	22168	GAGCTTAGAGCTTGATTCGAACTTGTCTGTGATAGTGGCCAAACCCAGCCGCAAGGCATCAG	2222
Qy	2220	AAACAAAGAGGCGCTGGGGCAACTATGCAATGTGCAAAAGGATTTGGCTCAGAGTTGTGGG	2279
Db	22228	AAACAAAGAGGCGCTGGGGCAACTATGCAATGTGCAAAAGGATTTGGCTCAGAGTTGTGGG	2228
Qy	2280	TAGAGAGTCCAAATCTGGGGGAGCTCAAAATTAATGTTCTTGGGTGATTCAAGTAACACACT	2339
Db	22288	TAGAGAGTCCAAATCTGGGGGAGCTCAAAATTAATGTTCTTGGGTGATTCAAGTAACACACT	2234
Qy	2340	CATGGCTTGTGTGCCATGAGTTAGGCAATACAGATGGAATGAAGTTGAAGTGGGAAAC	2399



Qy	8423	GTCGAATCTCATCTCCTACTAAAAATAAATAAAATTTACCAAGGATGTGGTGCATGCACAG	8482
Dd	20645	GTCAAACCACATCTCTACTAAAATTAC-AAAACATTAGCCAGGTGTGCTAGTGATGCATGCCCTG	20588
Qy	8483	TAGTCCCAAGCTGTTCAAGAAGCGTAGGGTGGAGAGATTGCTCGAGTGTGGAGGCCAAGAT	8542
Dd	20586	TAGTCCCAAGCTACTCAAGAGGCGTAACAAGAGAACGTTTGGAAACCAAGAGGCGAGGGT	20522
Qy	8543	TGCAGTGAACCGGTACCTGTGCCTTGGCAATCCAGCCTGGGGTGAACAATTGAGACCTGTCT	8602
Dd	20526	TGCAGTGAAGCCMAAGATTGCATCACTGCACATCCACAGTGGGGTGAACAAGTGAAGCATCCTATC	20467
Qy	8603	TCAAAAAACAAATTAATTAATAAATAAATAAATA 8634	
Dd	20466	TCAAAAATAAATAAATAAACAATAAATAAATA 20435	

Query	Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
Db	7652	62.1%	3.1*	4	45989	58	6
Db	21491	0	Pred. No. 2.6e-62				
Db	7712	0	Mismatches 318				
Db	21431	0	Indels 58				
Db	7772	0	Gaps 6				
Db	21371	0					
Db	7832	0					
Db	21315	0					
Db	7892	0					
Db	21296	0					
Db	7952	0					
Db	21238	0					
Db	8006	0					
Db	21178	0					
Db	8066	0					
Db	21118	0					

Oy	8126	CTCCCTCAGTTCCTCTTTTTCCTATTAAGATTTCCAGAGAGGAGATCTGAATGGCCCG	8185
Db	21059	GATAGCTCTTAGAACCCCAATCTCAGAGCTCATCTCCCTTAGTAAACATTTATTTCACG	210000
Oy	8186	TCCATATTTTCAGACCAACCACTTTAAAGTGTTGATTTCCAGCCTATGTAATTGGCTAC	8245
Db	20999	TAAATTAGCTTGAGTGCATTTTAGATCATCAGCTCTTAGAATAAGATTGTTTACCTCA	20940
Oy	8246	ATTAAATGGTTGGAACTCATCTATTACTTCATTCGACAAAGACAGATAGCTCTGGTTCT	8305
Db	20939	AATACTATGATCTAAATGCACTCAGACCTGAAATTTAACTGAATTTAAATTTTACTTTAA	20880
Oy	8306	CAAAATGGGCCC---CTGGCCACGGTGTGGTGGCTCATGCCCTATTAATCCCAACCTGG	8362
Db	20879	AATTTATTAACACATCTCTAGAGATGGGCATGGTGGCTCACCTGTATCTCCAGCACTTGG	20820
Oy	8363	GGAGGCCGAGGGGGGAGATCACTTAGTCCAGAGATTCTAGACCAAGCTGGGGCAACATG	8422
Db	20819	GGAGGCTGAGGAAAGCGGATCACTTAGGTGAGAGATTGAGACCAAGCTGGCCAACTGG	20760
Oy	8423	GTGAATCTCATCTCTACTTAAATAACAATAATTAGCCAGGTGTGGTGGCATGACCCAG	8482
Db	20759	GTCAAACTCATCTCTACTTAAATTAC-AAACATTAGCCAGGTGTGCTAGTGCATGCTGG	20701
Oy	8483	TAGTCCAGCTGTTTCAGAGAGGCTGAGGTGGAGAGATTGCTCGAGTGTGGAGGCGAGAT	8542
Db	20700	TAGTCCAGCTACTCAGAGAGGCTGAAAGCGAGAACTGTTTGAACCCAGAGAGCGAGGT	20641
Oy	8543	TGCAGTGAACCGTGACGTGTGCTCTGCAATCCAGCCTGGGTGAACGATTGAAACCTGTG	8602
Db	20640	TGCAGTGAAGCAAGATTGCACTGCACCTCCCAACCTGGGTGAACAGATGAGATCTTCATC	20581
Oy	8603	TCAAAAAACAATTAATTAATAATTAATAATA 8634	
Db	20580	TCAAAAAATTAATTAATTAACAATTAATAATA 20549	

```

RESULT 7
US-09-851-896-3
; Sequence 3, Application US/09651896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Preler
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: R1S-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match          3.1%; Score 305.8; DB 4; Length 70000;
Best Local Similarity 62.3%; Pred. No. 6,9e-62;
Matches 610; Conservative 0; Mismatches 312; Indels 57; Gaps 6;

QY 7650 AGTTTGTGTTGTGTTGCTGTGTTGAGATGAGAGTCACTCTGTCAACCGAGGCTGG 7709
      |||
Db 36060 AATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGAGAGGAGTCTGTGCTGTGTGCGCCAGGCTGG 36119

QY 7710 AGTGCATGGCGCAATCTTGGTTCACTGCAACCTCTGCTCTGGGTTCAAGGCAATCCTC 7769
      |||
Db 36120 AGTGCAGTGGCGCGAATCTCGGCTCACTGCAGACTCGGCTCCCGGGTTCAAGCAATCTC 36179

QY 7770 CTGGCTCAGCCCTCCCGGATAGCTGAGGAGCTACAGGCGGCAATACACCAATGCTGTGCTAATTT 7829
      |||
Db 36180 CTGGCTCAGCCCTCCCGGAGTGGAGACTACAGGCGGCGGCAACAGCGCCGGCTAATTT 36239

```

```

RESULT 8
US-09-539-333D-1
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marca
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguetelerc, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essiloux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30

```

```

PRIORITY APPLICATION NUMBER: US 60/126,903
PRIORITY FILING DATE: 1999-03-30
PRIORITY APPLICATION NUMBER: US 60/131,971
PRIORITY FILING DATE: 1999-04-30
PRIORITY APPLICATION NUMBER: US 60/132,065
PRIORITY FILING DATE: 1999-04-30
PRIORITY APPLICATION NUMBER: US 60/143,928
PRIORITY FILING DATE: 1999-07-14
PRIORITY APPLICATION NUMBER: US 60/145,915
PRIORITY FILING DATE: 1999-07-27
PRIORITY APPLICATION NUMBER: US 60/146,453
PRIORITY FILING DATE: 1999-07-29
PRIORITY APPLICATION NUMBER: US 60/146,452
PRIORITY FILING DATE: 1999-07-29
PRIORITY APPLICATION NUMBER: US 60/162,288
PRIORITY FILING DATE: 1999-10-28
PRIORITY APPLICATION NUMBER: US 09/416,384
PRIORITY FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 31..1107
OTHER INFORMATION: 5'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bb1s g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 64666..64812
OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc_feature
LOCATION: 65854..67854
OTHER INFORMATION: 3'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon
LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:

```

[illegible]



NAME/KEY: misc feature  
LOCATION: 247803..249803  
OTHER INFORMATION: 3'regulatory region  
NAME/KEY: allele  
LOCATION: 8316  
OTHER INFORMATION: 99-27943-150 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 21672  
OTHER INFORMATION: 99-27935-193 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 65485  
OTHER INFORMATION: 8-128-33 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 95396  
OTHER INFORMATION: 99-31960-363 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 107281  
OTHER INFORMATION: 99-24656-260 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 160640  
OTHER INFORMATION: 99-24639-163 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 160876  
OTHER INFORMATION: 99-24634-108 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 168974  
OTHER INFORMATION: 99-7652-162 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 170810  
OTHER INFORMATION: 99-16100-147 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 173358  
OTHER INFORMATION: 99-5862-167 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 189957  
OTHER INFORMATION: 99-5919-215 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 197163  
OTHER INFORMATION: 99-24658-410 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 200778  
OTHER INFORMATION: 8-303-235 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 202651  
OTHER INFORMATION: 8-300-221 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 202679  
OTHER INFORMATION: 8-300-193 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 203378  
OTHER INFORMATION: 8-299-128 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 204138  
OTHER INFORMATION: 8-296-213 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 204605  
OTHER INFORMATION: 8-252-190 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 204934  
OTHER INFORMATION: 99-24644-194 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 205206  
OTHER INFORMATION: 8-295-248 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 205329  
OTHER INFORMATION: 8-295-125 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 206064  
OTHER INFORMATION: 8-293-130 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 206545  
OTHER INFORMATION: 8-292-198 : polymorphic base A or G  
NAME/KEY: allele

LOCATION: 207313  
OTHER INFORMATION: 8-251-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208285  
OTHER INFORMATION: 8-289-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208960  
OTHER INFORMATION: 8-287-249 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 209123  
OTHER INFORMATION: 8-287-86 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 209631  
OTHER INFORMATION: 8-285-319 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210361  
OTHER INFORMATION: 8-283-278 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210463  
OTHER INFORMATION: 8-283-176 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210486  
OTHER INFORMATION: 8-283-153 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210583  
OTHER INFORMATION: 8-283-56 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 210879  
OTHER INFORMATION: 8-282-345 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210964  
OTHER INFORMATION: 8-282-260 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 210979  
OTHER INFORMATION: 8-282-245 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 211050  
OTHER INFORMATION: 8-282-174 : variable motif AAAG or GAAGGAAGGAAGGAAGA  
NAME/KEY: allele  
LOCATION: 211132  
OTHER INFORMATION: 8-282-92 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 211247  
OTHER INFORMATION: 8-281-367 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211315  
OTHER INFORMATION: 8-281-299 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211366  
OTHER INFORMATION: 8-281-248 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 212520  
OTHER INFORMATION: 8-279-197 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 212821  
OTHER INFORMATION: 8-278-289 : polymorphic base C or T  
NAME/KEY: allele

Query Match 3.0%; Score 305; DB 4; Length 319608;  
Best Local Similarity 59.7%; Pred. No. 2,2e-61;  
Matches 894; Conservative 0; Mismatches 454; Indels 149; Gaps 17;

QY 7173 AACACCTATATTAATAATACCAAGCCGCGGTGTGATCAAGCTTATATCTTACGA 7232  
DB 249487 AAGCTTTTAAACATCAATTTGAGACCGGCGGCTCATGCTTAATCCACGA 249546  
QY 7233 CTTTGGAGGCGTGAAGGCGGTGATGCTTGAAGCGGCTTGAAGCGGTTGAGACGA 7292  
DB 249547 CTTTGGAGGCGCAAGGTGGTGAATC-----AGGAGTTGATACCA 249587  
QY 7293 GCCTGGCAACATGCG-AAAACCTCATCTTCAAAAATATATAATAGTCGGGTG 7351  
DB 249588 GCCTGGCAACATGCGTGAAGAAACCCCGCTCTACTGTAATAATACAAAATTAAGCAGGTGAG 249647



OTHER INFORMATION: 5-131-395 : polymorphic base A or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106940  
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108106  
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108149  
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108308  
OTHER INFORMATION: 5-135-357 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108471  
OTHER INFORMATION: 5-136-174 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134134  
OTHER INFORMATION: 5-140-120 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134362  
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A  
FEATURE:  
NAME/KEY: allele  
LOCATION: 134374  
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA  
FEATURE:  
NAME/KEY: allele  
LOCATION: 146328  
OTHER INFORMATION: 5-143-84 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 146345  
OTHER INFORMATION: 5-143-101 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 150329  
OTHER INFORMATION: 5-145-24 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 160031  
OTHER INFORMATION: 5-148-352 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 72771..72817  
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30  
FEATURE:  
NAME/KEY: allele  
LOCATION: 72771..72817  
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51  
FEATURE:  
NAME/KEY: allele  
LOCATION: 86050..86096  
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31  
FEATURE:  
NAME/KEY: allele  
LOCATION: 86050..86096  
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52  
FEATURE:  
NAME/KEY: allele  
LOCATION: 90819..90865  
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49  
FEATURE:  
NAME/KEY: allele  
LOCATION: 90819..90865  
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70

FEATURE:  
NAME/KEY: allele  
LOCATION: 93690..93735  
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32  
FEATURE:  
NAME/KEY: allele  
LOCATION: 93690..93736  
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97099..97145  
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97099..97145  
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97130..97177  
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97130..97177  
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99075..99121  
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99075..99121  
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99094..99140  
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35  
FEATURE:  
NAME/KEY: allele  
LOCATION: 99094..99140  
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56  
FEATURE:  
NAME/KEY: allele  
LOCATION: 103783..103828  
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36  
FEATURE:  
NAME/KEY: allele  
LOCATION: 103783..103828  
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106918..106966  
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106918..106966  
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108084..108130  
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108084..108130  
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108127..108177  
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108127..108177  
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60  
FEATURE:



Query Match	3.0%;	Score 302;	DB 4;	Length 162450;
Best Local Similarity	61.1%;	Pred. NO. 8.3e-61;		
Matches 622;	Conservative	0;	Mismatches 315;	Indels 81; Gaps 5;

[illegible]

```

RESULT 11
US-09-851-896-3/c
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-DEPENDENT)
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

```

Query Match	3.0%;	Score 298.4;	DB 4;	Length 70000;
Best Local Similarity	63.9%;	Pred. No. 3.9e-60;		
Matches 626;	Conservative 0;	Mismatches 291;	Indels 63;	Gaps 9;

QY	7655	TGTTGTTGTGTGCTGTGTTGATGATGAGCTCAATC-TGTCAACCAAGGCTGAGATG	7713
Db	36982	TTTTTGTGTTTGTGTTTGTGTTTGTGAGTGAAGTTTCACTCTGTGTGCCAGGCTGTATG	36923
QY	7714	CAATGGCCGAATCTTGTGTTCACTGCAACTCTGCTCTCTGGGTTCAAGCAATCTCTCTGC	7773
Db	36922	CAGTGAACAATCTCTGGCTCACTGCAACCTCCGGCTC-CTGGGTTCAAGCAATCTCTCTGC	36864
QY	7774	CTCAGCCCTCCCGTATAGCTGGGAGCTACAGAGCATACCAACATAGCCCTGGCTAATTTTGT	7833
Db	36863	CTCAGCTCTCCGAGTAGCTGGGATTTACAGGCGTGGCCACCAAGCCTTGTCTAAT-----	36810
QY	7834	ATTTTGTGTTGGTTACAATGACTATATTTAATTTAATTTTGTATTTTGTATAGAGA	7893
Db	36809	-----TTTTGTATTTTGTATAGAGA	36789
QY	7894	TAGGTTTCAACATGTTGGCCAGGCTGTCTGAACTCTCGAATCTGAGATGATCCGCTG	7953
Db	36788	CGGGTTTACACACCTGTGCCAGGCTGTCTCAATCTCTGACTTCAAGATATACCCG	36729
QY	7954	CCTCGGCTCTCCCAAGTGTAGATTTACAGGATGAGCAACCAAGCTGTGCC-CTGTATA	8012
Db	36728	CCTGGGCTTCAAAAGTGTGGATTTACAGGTGTAGCCACCAAGCCGGCCAGCAATGA	36669
QY	8013	GGTTTATATCAATCTCTTGTCTTTCACAACTTGTGGCTGCAAGTGGACCATATGTT	8072
Db	36668	GGTTTGTGTTCTGGGCTGTGGAGCTTAATCTGAGCTATCTAGATATGAGACTAGAGAG	36609
QY	8073	CTCTCACTTCTCACTACTTCAATGATCTTTGAGTCTCAAGTTTCAACTGATACCTCCCTC	8132
Db	36608	GTTG-GGTACTCTGTGCACAAAGTGGCAACTGTGTCCCTGTGTCAAGGGTGGGACAGTG	36550
QY	8133	AGTTGCTCTTTTTTCTTAGTAAGATTTCCAGAGAGGAAATCTGAATGGCCCACTCATAT	8192
Db	36549	CCCT-CACCAAGGGGACAGAGCTGCTCATCTGTGCTGAATGATTTCAAAAAGGAGGC	36491
QY	8193	TTTCAGACCAACCACTTAAGTGTGATTTGGCCAGGCTATGTATGGTCACTATATG	8252
Db	36490	CTTACGCTTCCAGAGACAGCCTGGGACAGACACCAAGATTCAGGCCCTCTTTTCAACC	36431
QY	8253	GGTTGGGAATCATCAATTTACTTCAATGCACAAGCAGCATAGCTCTGTCTCAAAATA	8312
Db	36430	TCTCTGG-----CTGAGTGAAGCTGCAAAAGTCTGTATAGCTCT-CTTAGATATGA	36382
QY	8313	GGGCCCCGTGGGCCAGGTGTGTGTGCTCATGCTTATATTTCCAACTCTGTGGAGGCCGAG	8372
Db	36381	TGATTAACCTGGGCGGAGATGTGTGGCTCACGCTGTATATCCCGCACCTTGTGGAGGCCAG	36322



QY 8373 GGGGGGAGATCACTTGAAGTCCAGAGGTTCTAGACCAAGCTGGGCAACATGGTAATCTC 8432  
DB 36321 GAGGGGAGATCACTTGAAGTCCAGAGGTTCTAGACCAAGCTGGGCAACATGGTAATCTC 36262  
QY 8433 ATCTCTACTAAAAA--TACAAAAAATTAGCCAGGTGTGGTGGATCAGCAGATGTCCCA 8490  
DB 36261 GCTCTACTAAAAAACAACAAAAAATTAGCCAGGTGTGGTGGATCAGCAGATGTCCCA 36202  
QY 8491 GCTGTTCAGAGGCTTGAAGTGGGAGATGCTCTGAGTGGGAGGAGATTTGCACTGA 8550  
DB 36201 GCTACCGGGAGGCTGAGGAGGAGATGCTCTGAGTGGGAGGAGATTTGCACTGA 36142  
QY 8551 ACCGTGAGTGTGCTGTGCAATCCAGCTGGTGGAGAGATTTGAGACCTGTCTCAAAAA 8610  
DB 36141 GCGGAGATGCGCCCACTGCACTGAGCTGGGAGGAGAGATTTGAGACCTGTCTCAAAAA 36082  
QY 8611 CAATTAATTAATAATAATA 8630  
DB 36081 AAAAAAAAAAAAAAAAAA 36062

RESULT 12  
US-09-078-294-3/c  
; Sequence 3, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Deslree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 80595  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of HC-contig  
US-09-078-294-3

Query Match 3.0%; Score 296; DB 3; Length 80595;  
Best Local Similarity 62.9%; Pred. No. 1,5e-59;  
Matches 608; Conservative 0; Mismatches 290; Indels 68; Gaps 7;  
QY 7681 ATGAGTCTCACTGTGTCAACCCAGGCTGAGTGCATATGGGCAATCTGTGTCACTGCA 7740  
DB 39668 ATGAGTCTGTGTGTGT-TGCCAGGCTGGAGTGCATGTAC--ATCTGGGTTATGCA 39613  
QY 7741 CCTGTGCTCTGGGTTCAAGCATCTCTGCTCAAGCTTCCGTAATGCTGGAGCTAC 7800  
DB 39612 CCTCCACCTCTAGGTTCAAGCAATCTCTCCCTCAACCTCTGTAGTGGAGCTAC 39553  
QY 7801 AGGCGCATACCAACATGCGCTAATTTTGTATTTTGGTTACAATGTACTAT 7860  
DB 39552 AGGTGATGCAACCAACCTGGCTAATTTTGGC----- 39519  
QY 7861 TTAATTAATTAATTTTGTATTTTGTAGATAGAGTGGTTCAACATGTTGGCAGGCTG 7920  
DB 39518 -----TTTTAGTAGAGACAGGTTTCAACATGTTGGCAGGCTG 39479  
QY 7921 GCTCTCAATCTGACCTCAAGTATCCGCTGCTGCTGCCCTCCAAAGTGTAGATTA 7980  
DB 39478 GCTCTCAATCTGACCTCAAGTATCCGCTGCTGCTGCCCTCCAAAGTGTAGATTA 39419  
QY 7981 CAGGCAATGACCAACGCGCTGGCCCTCATATA-----GGTTTATCTATCTCTTGTCT 8035  
DB 39418 CAGGCAATGACCAACGCGCTGGCCCTCATATA-----GGTTTATCTATCTCTTGTCT 39359  
QY 8036 TCTTCAAACTTGTGCTGACATGCAATGCTCTCTCACTTCTCACTTCA 8095  
DB 39358 TTTTAAATTTTATTAATAGTTGTAAATAAATAAATCTTCAATGACACTAC 39299

QY 8096 TGAATCTTCAAGTCTCAAGTTCAACTGATACCTCCCTCAGTGTCTTTTCTCTAGTAAG 8155  
DB 39298 TACTTGTTCATTAATCTTATTTTATATATCAACACTATCTCCAAACATTTTCAATAGTTAT 39239  
QY 8156 ATTTCCAGAGAGGAATCTGAATGGCCAGTCCATATTTTTCAGACCAACCAATTAAG 8215  
DB 39238 TCTCTTGAAGGCAATCATTTATTTTCACTTCAATATGAGAAAGACAGCTGAGAGAC 39179  
QY 8216 TGGTGAATGCAAGCTATATATTTGGCTATCAATTAATGGTTGGAACTCATCTTACTT 8275  
DB 39178 TAAATCATTTAAGGCGCTTCAAGTAAATGAGCAGATTTGGATTTGACTCATCAAGTCTG 39119  
QY 8276 CATTCGACAAAGCAG-----CATPAGCTGTGGTTCTCAAAATPAGGCGCCCTGGCC 8325  
DB 39118 TCAATTAAGAGACAGATTTTAAGGCCCATATTTGTCTTCAATTAATTAAGACATTTGACC 39059  
QY 8326 AGGTGTGTGCTCATATGCTTATATCCCAACATGTGGAGGCGGAGGGGAGATAC 8385  
DB 39058 AGCGGTGTGTGCTCATATGCTTATATCCCAACATGTGGAGGCGGAGGGGAGATAC 39000  
QY 8386 TTGATCCAGAGTCTTGAACCAAGCTGGGCAACATGTGAATCTCATCTCTAATAA 8445  
DB 38999 -TGAGGTCAAGAGTTCAGACCAAGCTGGGCAACATGTGAATCTCATCTCTAATAA 38941  
QY 8446 ATACAAAAAATTAAGCAGGTGTGTGTCATGACAGATGTGCAATGTGCTGTGAGAGGCT 8505  
DB 38940 ATAC-AGAAATTAAGCAGGTGTGTGTCATGACAGATGTGCAATGTGCTGTGAGAGGCT 38882  
QY 8506 GAGGTGGAGAGATTCCTCAAGTGTGGAGGAGAGATGTGCAATGTGCTGTGAGAGGCT 8565  
DB 38881 GAGGAGAGAGATTCCTCAAGTGTGGAGGAGAGATGTGCAATGTGCTGTGAGAGGCT 38822  
QY 8566 CTGCAATCCAGCTGTGGTGTGACAGATTTGAGACCTGTCTCAAAAAAATAATAATA 8625  
DB 38821 CTGTACTTACGCTGTGGTGTGACAGAGATTTGAGACCTGTCTCAAAAAAATAATAATA 38762  
QY 8626 AATAA 8631  
DB 38761 AATAA 38756

RESULT 13  
US-08-370-975B-6/c  
; Sequence 6, Application US/08370975B  
; Patent No. 5622851  
; GENERAL INFORMATION:  
; APPLICANT: Maley, Frank  
; APPLICANT: Maley, Gladys F.  
; APPLICANT: Weiner, Karen X.B.  
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/370,975B  
; FILING DATE: 10-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Timian, Susan J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 20894/80  
; TELECOMMUNICATION INFORMATION:



QY	7830	TGTAATTTTTTTGGGTACAAGTACTAATTATAATTTATTTGA	7889
Dd	17311	-----TAATCTTGATTTTAA	17292
QY	7890	GAGATAGGGTTTCACATGTTGGCAGGCTGATCTCGAACTCTGACCTCAGTGATCCG	7949
Dd	17291	GAACACAGGGTTTCGTATGTTGGCAGGCTGATCTCGAACTCTGACCTCAGTGATCCA	17232
QY	7950	CTTGCCCTCGGCTTCCCAGAGTCTAGAGTTACAGGCATGAGCAACACGCCCTGCCCTC	8009
Dd	17231	CCTGCCCCGCGCTCAAAAAGTCTGGAGTTACAGGCATGAGCACCGGCCCGCCAGTTC	17172
QY	8010	AATAGTTTATCATATCTCTTGCTCTTCAACAATTTGGCTGACGATGACCATCAT	8059
Dd	17171	ACTAACCTTAAAGAAATCCATCTGTTCAATGATCTGAGTCCAGATTCATCTGTAACCT	17112
QY	8070	GTTCTCTCACTTCTGCACTA-----CTTCATGATCTTTCAGTCTCACTTC	8115
Dd	17111	GAATCTGATGATCTAGCATTAACGTCATCTTCTCCAGGCTTTCTTTGGGTGAT	17052
QY	8116	CAACTGATACCTCCCTCAGTGTCTTTTCTTAGTAAGATTTCCAGAGGGAATCTG	8175
Dd	17051	CTGGGGCTTCATGATCACTTCTCTGATGGGTAAATCTCTATCCAAATCTTAGG	16992
QY	8176	AATGGCCCAATCCATATTTTAGACACACACACATTAAGTGTGATGTCACACCTATG	8235
Dd	16991	AGCAGAAATGTTTTCAGATTTTCAGATTTTAAATGTGTGATTAATTAACAGGTGAGCA	16932
QY	8236	TATGGCTACATTAATGGGTGGGAACATCATTTACTTATGSCAACAGCAGCATAG	8295
Dd	16931	TCCCAATTCCAAATTTTGAATTCAGAGTCTCATGAGCATTTCTTCAAGTGTATG	16872
QY	8296	CTCTGTTCTCAAA-ATAAGGCCCTCGGGCAGGTGTGTGATCATGCTATTAATCCA	8354
Dd	16871	CTGATGTGCAAAAAGSTTTCAATTTTGGGCGAGGTGAGTATCATGCTGTATATCCA	16812
QY	8355	ACACTGTGGAGGCGCCAGGGGGGCGAGATCATTTGATTCAGAGTTTAAACCAGCTGG	8414
Dd	16811	GCACCTTTGGGAGGCGCAAGTGGGGAGATCGTTTAGCTCAAGATTCAGAAGCAGCTTAG	16752
QY	8415	GCAACATGATGTAATTCATCTCTACTTAATAATAAAAAATTAGCCAGGTGTGTGGCA	8474
Dd	16751	GCAACATGATGACAGCTCATTTCTACTGTAATAATAAAAAATTAGCCAGGTATGTGGCG	16692
QY	8475	TGCACAGTAGTCCAGCTGTTTCAGAGGCTGAGGTGGAGATTTGCTCGAGTGTGGAG	8534
Dd	16691	TGCACCTGATGCTCTAGTACTTTGGAGGCTGAGGTGGAGAGATGCTGTAGCCACAGAG	16632
QY	8535	GCAGAGATTCAGTGAAACCGTGAATGCTCTCGAATCCAGCTGGGTGACAGATTGAG	8594
Dd	16631	GTCAAGGCTGAGTGAGGCTGTGTTTGCACCACTGATCTCAGTCTGGGCAACAGCAGAG	16572
QY	8595	ACCTGTCTCAAAAAACCAATATAA	8618
Dd	16571	ACCCCATCTCAGAAAAAATAAAAAA	16548
 RESULT 15 US-09-491-356C-1/C Sequence 1, Application US/09491356C Patent No. 6566061 GENERAL INFORMATION: APPLICANT: Philibert, Robert A. APPLICANT: Ghans, Edward I. TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XOL3 FILE REFERENCE: 9465.6US11 CURRENT FILING DATE: US/09/491.356C PRIORITY FILING DATE: 2000-01-26 PRIOR APPLICATION NUMBER: PCT/US99/09365 PRIOR FILING DATE: 1999-04-29 PRIOR APPLICATION NUMBER: 60/083,465 PRIOR FILING DATE: 1998-04-29			

	NUMBER OF SEQ ID NOS: 24	
:	SOFTWARE: PatentIn version 3.1	
:	SEQ ID NO 1	
:	LENGTH: 55298	
:	TYPE: DNA	
:	ORGANISM: Homo sapiens	
:	FEATURE:	
:	NAME/KEY: misc feature	
:	LOCATION: (485)..(485)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (838)..(838)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (16728)..(16728)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (22750)..(22750)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (22756)..(22756)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (28519)..(28519)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (44804)..(44804)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (45002)..(45002)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (54049)..(54049)	
:	OTHER INFORMATION: n is not determined	
:	NAME/KEY: misc feature	
:	LOCATION: (54226)..(54226)	
:	OTHER INFORMATION: n is not determined	
:	US-09-491-356C-1	

  

Query Match	2.9%; Score 294.8; DB 4;	
Best Local Similarity	61.3%; Pred. No. 2.5e-59;	
Matches 601; Conservative 2; Mismatches 325; Indels 53; Gaps 6;		

  

Dy	7641 CTGACTCACAGATTGTGTGGTGTGTCGTGTGTGAATGAGAGTCTCACTCTGTAC	7700
Db	16910 CTGAGTTCAAGCTTGAGATTTTTTTTTTTTTTTTTGAACCGAAGTCTGCCTGTAC	1685
Dy	7701 CCAGGCTGAGGTGCAATGGCGCAATCTTGCTTCACTGCAACTCTGCTCTGGGTCAA	7760
Db	16850 CCAGGCTGAGGTGCAATGGCGCAACTCTGCTTCACTGCAAACTCCGCTCCGGGTTCAC	1679
Dy	7761 GCATCTCTCGCTCAAGCTCCCGATAATCTGGAACTACAGGGCATACACAGGCT	7820
Db	16790 GCCATTCTCTGCTCAAGCTCCCGAATCTGGAAATCAAGGCGCTGCCACACGCC	1673
Dy	7821 GGCTAATTTTGTATTTTGTGGGTACATGTAATAATTAATTTTGTGA	7880
Db	16730 AGNMAATT-----TTTTGTA	1671
Dy	7881 TTTTAGTAGAATAGGTTTCAACATGTGGCCAGGCTGCTCGAATCTCGAAGCTCA	7940
Db	16713 TTTTAGTAGAAGAGCGGGGTTTCATCGGTAAACAGAGATGTCGATCTCTGACCTC-	1665
Dy	7941 GGTGATCCGCTGCTCGGCTCCCAAGTGTAGAATTAACGGCAAGAACCAACAGCC	8000
Db	16654 -GGATCCGCCCACTCGGCTCCCAAGTGTAGGATTAACGAGAGAACCCACCGGCC	1659
Dy	8001 TGCCCTCATAGATTATTAATCTCTTGCTTCTTCAACAATTGGCTTGACGTG	8060
Db	16595 CGGCATGCTTGAAGTTAGCCTATGATTTCATATCATCTCAAGAACATGACGGAATT-	1653
Dy	8061 GACCATCATGTTCTTCACATTTCTCATTAATTCAATGATCTTTCAGTCTCAGTTCAACT	8120

Db 16536 -CTTGAAGTCTCAGGCTCTGTTTCCATCTTAGAAAGTATGGGAATTGACAGAAATG 16478  
QY 8121 GATACCTCCCTCAGTTCCTTTTCTCTAGTAAGATTCCAGAGGAATCTGAATG 8180  
Db 16477 ATTTCCAGGAATTCAT--TCAACATTCACAGTGGGTTCCAGTGTCTTAGGGCCCAT 16420  
QY 8181 CCCAGTCCATATTTTCAGACCACACACATTTAAGTGTGATTCAGCCCTATGTATG 8240  
Db 16419 CAGATCTAGGATTAAGTACAGGACAGAGGCTGCAAGATCCAGGGTACATAACTATGGA 16360  
QY 8241 GCTACATTAATGGTGGGAATCATCTTACTTCAATGCT--ACAAGCAGATAGCT 8297  
Db 16359 CCATGACCAAGATGATCCAGCCCCCTTAAACACTTTCAGTCCCTGCACTTTCAGAGAA 16300  
QY 8298 CTGGTTCCTAAATATAGGCCCCCTGAGCCAGGTGTGAGGCTCATGCTTATATCCCAAC 8357  
Db 16299 CAAGGTTAATATCTTCAAGACTCAAGCTGGGACAGGAGGCTCATGCTGTATTCAGCA 16240  
QY 8358 CTGTGGAGGCCGAGAGGGGGCAGATCACTTGAATCCAGAGTTCTAGACCAAGCCTGGCA 8417  
Db 16239 CTTTGGGAGGCCGAGACAGACGAGTCAATTGAGGTAGAGTTCAAGACCAAGCCTGACCA 16180  
QY 8418 ACATGGTGAATCTCATCTCTATAAATAACAAAAATTAGCCAGGTGTGGCATGC 8477  
Db 16179 GCATGGTCAAAACCCCATCTCTATGAAAAATAC--AAAAATTAGCCAGGTGTGGTACAC 16121  
QY 8478 ACCAGTAGTCCAGCTGTTCAGGAGGCTGAGGTGGAGGATTTGCTGAGTGTGGGAGGCA 8537  
Db 16120 ACCGTATATCCAGCTACTCAAGAGGTTGAGGACAGAAATCACTTGAACCCAGAGAGTG 16061  
QY 8538 GAGATTGCACTGAACCGTGAATGTGCTCTGCAATCCAGCCTGGGTGACAGATTGAGACC 8597  
Db 16060 GAGGTGCAAGTGAAGTGAATGGGCGACCTGCACTCCAGCTTGGGTGACAGATGAGACT 16001  
QY 8598 CTGCTCAAAAAACAATATA 8618  
Db 16000 CTGACTCCAAAAAATAAAAA 15980

Search completed: December 16, 2003, 05:47:11  
Job time : 467.684 sec8



Db 181 CCATTGCCCAACACCCAGAAAAGACATGCTCCCTGCATATGGGAGAGAGTATATAG 240  
Qy 241 CATTTGCTGTAGAGAGGATGGCAATGGCCAGGCTGCCAAGACTCAGAGGTCCACCTT 300  
Db 241 CATTTGCTGTAGAGAGGATGGCAATGGCCAGGCTGCCAAGACTCAGAGGTCCACCTT 300  
Qy 301 GCCCACTGACCTTATGAGAGGAGATATGTTTCAAGACATTTTCAATTCGTATAGT 360  
Db 301 GCCCACTGACCTTATGAGAGGAGATATGTTTCAAGACATTTTCAATTCGTATAGT 360  
Qy 361 GAGAGCATTGAGCTGATGAGAGGCTGTGATCATGTTTCCAGAGGTTCCGAAATG 420  
Db 361 GAGAGCATTGAGCTGATGAGAGGCTGTGATCATGTTTCCAGAGGTTCCGAAATG 420  
Qy 421 TGTGTTTTCTGTGGAAGAACTTTCAGAGATGAAAAGGATCTGAGACTTTTGGTA 480  
Db 421 TGTGTTTTCTGTGGAAGAACTTTCAGAGATGAAAAGGATCTGAGACTTTTGGTA 480  
Qy 481 AGATTATATATGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTTTTCA 540  
Db 481 AGATTATATATGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTTTTCA 540  
Qy 541 GCCTTGCTAGGAGAGGAGGCTGTGAACTTTCATCTGTGCTGAGCTGATCTGCCA 600  
Db 541 GCCTTGCTAGGAGAGGAGGCTGTGAACTTTCATCTGTGCTGAGCTGATCTGCCA 600  
Qy 601 TAATTTCTTTTCTGATCTGATGAGCAATCTCAACAGAAATGGGGAGCTTTGGAGT 660  
Db 601 TAATTTCTTTTCTGATCTGATGAGCAATCTCAACAGAAATGGGGAGCTTTGGAGT 660  
Qy 661 GCCCAGAAATGATGAGATGAGAGGATGAGATGAGATGAGATGAGATGAGATGAG 720  
Db 661 GCCCAGAAATGATGAGATGAGAGGATGAGATGAGATGAGATGAGATGAGATGAG 720  
Qy 721 ACACTGAGAGAGAAACAGAGCTGTGATGATGAGAGGCTGTGATGATGATGAT 780  
Db 721 ACACTGAGAGAGAAACAGAGCTGTGATGATGAGAGGCTGTGATGATGATGAT 780  
Qy 781 ATGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 781 ATGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Qy 841 ATGCTGAGAAATTTGCTGAGCACTGAGCACTGAGCACTGAGCACTGAGCACTT 900  
Db 841 ATGCTGAGAAATTTGCTGAGCACTGAGCACTGAGCACTGAGCACTGAGCACTT 900  
Qy 901 CATTTTCTTCAATCAAAATTTTCTGAGAACTTATGATGAGGCTATTTTATG 960  
Db 901 CATTTTCTTCAATCAAAATTTTCTGAGAACTTATGATGAGGCTATTTTATG 960  
Qy 961 GAGTCTGAGATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Db 961 GAGTCTGAGATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Qy 1021 TCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
Db 1021 TCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
Qy 1081 AAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Db 1081 AAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
Qy 1141 AAGCAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
Db 1141 AAGCAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
Qy 1201 GGTATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
Db 1201 GGTATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
Qy 1261 TGTGCTGCTCTGAGAGAGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320  
Db 1261 TGTGCTGCTCTGAGAGAGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320

Qy 1321 ACTTCAGGCACTTTTGGAAAAGCCCTTGGCAGTCAAGGAAAGGCGAGAGGCTGGCAGT 1380  
Db 1321 ACTTCAGGCACTTTTGGAAAAGCCCTTGGCAGTCAAGGAAAGGCGAGAGGCTGGCAGT 1380  
Qy 1381 GGGCTTGGACATTTGACAAACAGTGAACATTTATGTTCCAGAGCTCACTAGCCCAAG 1440  
Db 1381 GGGCTTGGACATTTGACAAACAGTGAACATTTATGTTCCAGAGCTCACTAGCCCAAG 1440  
Qy 1441 AAAAGTGAAGAGCTTGGGCAATGCCAGAAAAGCCCTGATGAGCTTGGAAAAAGCTG 1500  
Db 1441 AAAAGTGAAGAGCTTGGGCAATGCCAGAAAAGCCCTGATGAGCTTGGAAAAAGCTG 1500  
Qy 1501 TTCTCTGAGATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
Db 1501 TTCTCTGAGATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
Qy 1561 TGTGCTCTGAGCTGCTTGAAGAGAGGCTTGAACAGTGAACAGCAATATTAGTTCCCT 1620  
Db 1561 TGTGCTCTGAGCTGCTTGAAGAGAGGCTTGAACAGTGAACAGCAATATTAGTTCCCT 1620  
Qy 1621 CTTTCTGACCTCCCACTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
Db 1621 CTTTCTGACCTCCCACTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
Qy 1681 ACTCACTTCCCTGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
Db 1681 ACTCACTTCCCTGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
Qy 1741 CGTCAAGAGTGAAGCCAGATTAAGTCTGTAAGGAGGCTTGGGGGAGCTG 1800  
Db 1741 CGTCAAGAGTGAAGCCAGATTAAGTCTGTAAGGAGGCTTGGGGGAGCTG 1800  
Qy 1801 CCGGAAAAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
Db 1801 CCGGAAAAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
Qy 1861 TGTATCAGCAATATGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1920  
Db 1861 TGTATCAGCAATATGAGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1920  
Qy 1921 TGGGGTGTGCTGCTGCACTATATAGGAGGATTAAGTCTGATGATGATGATGAT 1980  
Db 1921 TGGGGTGTGCTGCTGCACTATATAGGAGGATTAAGTCTGATGATGATGATGAT 1980  
Qy 1981 TGTGCTCTGAGCTCAGCATGCTGAGGATGATGATGATGATGATGATGATGATGAT 2040  
Db 1981 TGTGCTCTGAGCTCAGCATGCTGAGGATGATGATGATGATGATGATGATGATGAT 2040  
Qy 2041 GAGGCTATGCTGCTGCTGCACTTGAAGGAGGATGATGATGATGATGATGATGAT 2100  
Db 2041 GAGGCTATGCTGCTGCTGCACTTGAAGGAGGATGATGATGATGATGATGATGAT 2100  
Qy 2101 TAGAAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
Db 2101 TAGAAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
Qy 2161 CAAATTTGGGCTGAGGCTGATGAGAAAGGATGATGATGATGATGATGATGATGAT 2220  
Db 2161 CAAATTTGGGCTGAGGCTGATGAGAAAGGATGATGATGATGATGATGATGATGAT 2220  
Qy 2221 TTGTTCTCCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
Db 2221 TTGTTCTCCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
Qy 2281 CAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Db 2281 CAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Qy 2341 TGTATTTATCACTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTACTTTGCAATCTCCT 2400  
Db 2341 TGTATTTATCACTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTACTTTGCAATCTCCT 2400





[illegible]

QY	5701	CATCTCTTCCACCTTGGACCTCACTCTGACCTCTGGGCTCTCTTGTGTTCTCTCAACC	5768
Db	5701	CATCTCTTCCACCTTGGACCTCACTCTGACCTCTGGGCTCTCTTGTGTTCTCTCAACC	5766
QY	5761	ATTTCCTCTCTCAGAGCCATGGACACAGGAATGGAGACTGGCCATGGAAACAGTACCACA	5820
Db	5761	ATTTCCTCTCTCAGAGCCATGGACACAGGAATGGAGACTGGCCATGGAAACAGTACCACA	5820
QY	5821	TGGCCCTGAGTATATGCGCTGTAGCCCAACATCTAGTCTGTCTGCACTGACGTCTGCAA	5880
Db	5821	TGGCCCTGAGTATATGCGCTGTAGCCCAACATCTAGTCTGTCTGCACTGACGTCTGCAA	5880
QY	5881	CCATGTGTGCCACCTATGCTTCAGTGTGTAGAGATGCCCCCAATCTCCCCCAATGTGCTCT	5940
Db	5881	CCATGTGTGCCACCTATGCTTCAGTGTGTAGAGATGCCCCCAATCTCCCCCAATGTGCTCT	5940
QY	5941	CACATCTCTTCTTACTGTATCTGCCATCTTGAACAATTTCTCATATGTCTCATCTGCT	6000
Db	5941	CACATCTCTTCTTACTGTATCTGCCATCTTGAACAATTTCTCATATGTCTCATCTGCT	6000
QY	6001	CACCTAATTTTGTCCCTCTGTGCCCATCTTCTGCAATGCCCTTCTGCAATCCCTATCTCT	6060
Db	6001	CACCTAATTTTGTCCCTCTGTGCCCATCTTCTGCAATGCCCTTCTGCAATCCCTATCTCT	6060
QY	6061	GAGGCATATTTTCTCAATCTTGTCTGTCAACGGCCCAAGCCCTTAACCTATCTACCTGCT	6120
Db	6061	GAGGCATATTTTCTCAATCTTGTCTGTCAACGGCCCAAGCCCTTAACCTATCTACCTGCT	6120
QY	6121	ACCAATCTAATCCCAATGGCTGTGCCCCCTGTGGACATCTCTGGGGCCCTATGACCTCTGT	6180
Db	6121	ACCAATCTAATCCCAATGGCTGTGCCCCCTGTGGACATCTCTGGGGCCCTATGACCTCTGT	6180
QY	6181	GTTTCCTCTTGTCTCAATGCCCTCTGTAGAGCCCTCTGAGCCCTCTGCCCTGACCTCTTA	6240
Db	6181	GTTTCCTCTTGTCTCAATGCCCTCTGTAGAGCCCTCTGAGCCCTCTGCCCTGACCTCTTA	6240
QY	6241	TGTGTCTCTGTACTCCTTGTGCTTCCCTCTCTTGTCTTGCATATCTTTCAGAGTCTCTGGCT	6300
Db	6241	TGTGTCTCTGTACTCCTTGTGCTTCCCTCTCTTGTCTTGCATATCTTTCAGAGTCTCTGGCT	6300
QY	6301	CCCCCTGATTTATTCCTCAGAACTCCATCTTGTTCAGGTTCCCGGTTCCATATGCCAGAC	6360
Db	6301	CCCCCTGATTTATTCCTCAGAACTCCATCTTGTTCAGGTTCCCGGTTCCATATGCCAGAC	6360
QY	6361	CCCTGTGGCATGTGACCTGCTGGGGATAGATGTTCTCATTTGCTGAGAACACAGCTGAGAG	6420
Db	6361	CCCTGTGGCATGTGACCTGCTGGGGATAGATGTTCTCATTTGCTGAGAACACAGCTGAGAG	6420
QY	6421	TGTTGGGTACTTTAGACTTTTAAAGAGGCTTGCTTCACTAGCCCTGTGAGGTTTCTCTCTG	6480
Db	6421	TGTTGGGTACTTTAGACTTTTAAAGAGGCTTGCTTCACTAGCCCTGTGAGGTTTCTCTCTG	6480
QY	6481	AGTATGCCAATGTAGATACCCCTCCCTTGAACCCGTGGCATCAATTGGTGTAAAAAGCATCTTA	6540
Db	6481	AGTATGCCAATGTAGATACCCCTCCCTTGAACCCGTGGCATCAATTGGTGTAAAAAGCATCTTA	6540
QY	6541	TAAATACCTAGGGCTGTCTGAGTTCACTGACGGCAGTAAATAGTATGATCTGACAGTTGAG	6600
Db	6541	TAAATACCTAGGGCTGTCTGAGTTCACTGACGGCAGTAAATAGTATGATCTGACAGTTGAG	6600
QY	6601	AATATCCCAAGAGAGTGTAGCAACATCACTCACTCAACCTGAGATATATGTATTAATTA	6660
Db	6601	AATATCCCAAGAGAGTGTAGCAACATCACTCACTCAACCTGAGATATATGTATTAATTA	6660
QY	6661	GGACAGTGTATAGATATAAAAATCGTAAAAATTTTTTTCACAAAAATTTTTTGGCT	6720
Db	6661	GGACAGTGTATAGATATAAAAATCGTAAAAATTTTTTTCACAAAAATTTTTTGGCT	6720
QY	6721	CTTGAACCTTGGAAATTTTGAACAAGTTATGACATCAAGTCTGTGAAAAAATATACATCA	6780
Db	6721	CTTGAACCTTGGAAATTTTGAACAAGTTATGACATCAAGTCTGTGAAAAAATATACATCA	6780



QY 6781 CCACATGAGAGCAAAATCTCCACAGAGATTGACACTATATTAAGAAATACAGCTAA 6840  
DB 6781 CCACATGAGAGCAAAATCTCCACAGAGATTGACACTATATTAAGAAATACAGCTAA 6840  
QY 6841 GATGAAACACACCTGTAGTAAATACAACTTAACTGAGAACATAGCCATAGTA 6900  
DB 6841 GATGAAACACACCTGTAGTAAATACAACTTAACTGAGAACATAGCCATAGTA 6900  
QY 6901 GAAACATATATATGTAAGAGAACACACAGCCATGTGTGGAGCCCATTTGGAGAACACAC 6960  
DB 6901 GAAACATATATATGTAAGAGAACACACAGCCATGTGTGGAGCCCATTTGGAGAACACAC 6960  
QY 6961 GACAAAGTAAATGACAGAAAGAGAGAGAGTGAAGATTGTGAAACAGGGCCAC 7020  
DB 6961 GACAAAGTAAATGACAGAAAGAGAGAGAGTGAAGATTGTGAAACAGGGCCAC 7020  
QY 7021 AGGAAACACACAGAAATAGAGAGAACACCAAGCCATCTAGAGATCACAGAACTTCATGG 7080  
DB 7021 AGGAAACACACAGAAATAGAGAGAACACCAAGCCATCTAGAGATCACAGAACTTCATGG 7080  
QY 7081 CCATGTGGCATATATGAAATGCTACTGAACTCTTAATTAATAATGTCTATGTATTC 7140  
DB 7081 CCATGTGGCATATATGAAATGCTACTGAACTCTTAATTAATAATGTCTATGTATTC 7140  
QY 7141 ATAGCTGTAGAGAGAGCCACACAGATGAGAGAAACCTTATATTAATAATACCGAGCC 7200  
DB 7141 ATAGCTGTAGAGAGAGCCACACAGATGAGAGAAACCTTATATTAATAATACCGAGCC 7200  
QY 7201 GGGCGTGTAGAGTCAAGCCCTGTATCTTACACTTTGGAGGCTGAGGCAAGTGAATTC 7260  
DB 7201 GGGCGTGTAGAGTCAAGCCCTGTATCTTACACTTTGGAGGCTGAGGCAAGTGAATTC 7260  
QY 7261 TTGAGGCGCTTGAAGCTTGAAGATTGAGACACAGCTGGGCAATAGGCAAAACCTCATCT 7320  
DB 7261 TTGAGGCGCTTGAAGCTTGAAGATTGAGACACAGCTGGGCAATAGGCAAAACCTCATCT 7320  
QY 7321 CTACAAATAATATTAATAATTAAGTGGGTGTAGTGTGCTTATATAGTCCATCTACT 7380  
DB 7321 CTACAAATAATATTAATAATTAAGTGGGTGTAGTGTGCTTATATAGTCCATCTACT 7380  
QY 7381 CAGAGGCTGAGCCCGGAAAGGTGAGGCTTCAAGTGAAGCCGTGATCTGTACTACCTCCA 7440  
DB 7381 CAGAGGCTGAGCCCGGAAAGGTGAGGCTTCAAGTGAAGCCGTGATCTGTACTACCTCCA 7440  
QY 7441 GCGTGGGTGACAGAGTGAAGCATGTCTCAAAAACAAAACAAAACAAAACAAAACAAA 7500  
DB 7441 GCGTGGGTGACAGAGTGAAGCATGTCTCAAAAACAAAACAAAACAAAACAAAACAAA 7500  
QY 7501 CAAACAAACAAACAAACAAACCCATATATATATATATATATATATATATATATATAT 7560  
DB 7501 CAAACAAACAAACAAACAAACCCATATATATATATATATATATATATATATATATAT 7560  
QY 7561 CTATTTTGTAAATATACCAACATGACCCAGCTACAGATGGGAGTCCCTCCCTCTC 7620  
DB 7561 CTATTTTGTAAATATACCAACATGACCCAGCTACAGATGGGAGTCCCTCCCTCTC 7620  
QY 7621 ACTGTAAATTTTCTTCTCTGACTCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTG 7680  
DB 7621 ACTGTAAATTTTCTTCTCTGACTCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTG 7680  
QY 7681 ATGAGATCTACCTGTCAACCCAGGCTGAGTGAATGGGCAATCTGTGTTCACTGCA 7740  
DB 7681 ATGAGATCTACCTGTGTCAACCCAGGCTGAGTGAATGGGCAATCTGTGTTCACTGCA 7740  
QY 7741 CCTGTGCTCTGTGGTTCAGAGCATCTCTGCTCAGCTCCCGTATAGTGGAGCTAC 7800  
DB 7741 CCTGTGCTCTGTGGTTCAGAGCATCTCTGCTCAGCTCCCGTATAGTGGAGCTAC 7800  
QY 7801 AGGCGCATACACCATAGCTGCTGCTAATTTTGTATTTTGTGTGTGTGTGTGTGTGTGT 7860  
DB 7801 AGGCGCATACACCATAGCTGCTGCTAATTTTGTATTTTGTGTGTGTGTGTGTGTGTGT 7860  
QY 7861 TTATTAATTAATTTTGTATTTTGTATTTTGTATGAGATAGGCTTCACATGTTGGCAGGCTG 7920

DB 7861 TTATTAATTAATTTTGTATTTTGTATTTTGTATGAGATAGGCTTCACATGTTGGCAGGCTG 7920  
QY 7921 GTCTGCAATCTGACCTGAGGTATCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7980  
DB 7921 GTCTGCAATCTGACCTGAGGTATCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7980  
QY 7981 CAGGCAATGAGCAACCGCTGGCCCTCATAGTGTATTTATCATTTCTTGTCTTCTTCTT 8040  
DB 7981 CAGGCAATGAGCAACCGCTGGCCCTCATAGTGTATTTATCATTTCTTGTCTTCTTCTT 8040  
QY 8041 ACACTTTGGCTTGCACGTGACACATATGTTCTCTCACTTCTCACTTCTCACTTCACTATC 8100  
DB 8041 ACACTTTGGCTTGCACGTGACACATATGTTCTCTCACTTCTCACTTCTCACTTCACTATC 8100  
QY 8101 TTTCACTCACTGCTCACTGATACCTCCCTCATAGTGTCTTCTTCTTCTTCTTCTTCTTCT 8160  
DB 8101 TTTCACTCACTGCTCACTGATACCTCCCTCATAGTGTCTTCTTCTTCTTCTTCTTCTTCT 8160  
QY 8161 CAGAGGGAATCTGAAATGGCCCAATGCAATTTTTCAGACCAACCACTTAATAGGCTT 8220  
DB 8161 CAGAGGGAATCTGAAATGGCCCAATGCAATTTTTCAGACCAACCACTTAATAGGCTT 8220  
QY 8221 GATTGCACTATATGATTTGGCTACATTAATGGGTTGGAACTCATTTACTTCACTT 8280  
DB 8221 GATTGCACTATATGATTTGGCTACATTAATGGGTTGGAACTCATTTACTTCACTT 8280  
QY 8281 CACAAAGCAGATAGCTGTGTTCTCAAAATAGGCCCCCTGGGCGAGTGTGGCTCA 8340  
DB 8281 CACAAAGCAGATAGCTGTGTTCTCAAAATAGGCCCCCTGGGCGAGTGTGGCTCA 8340  
QY 8341 TGGCTATATTTCCCAACATGTGTGGAGCCGAGGGGGGAGATCACTTGAATCCAGAGTT 8400  
DB 8341 TGGCTATATTTCCCAACATGTGTGGAGCCGAGGGGGGAGATCACTTGAATCCAGAGTT 8400  
QY 8401 CTAGACCAAGCTGGGCAATGATGTAATCTCATCTCTAATAATAATAATAATAATAATA 8460  
DB 8401 CTAGACCAAGCTGGGCAATGATGTAATCTCATCTCTAATAATAATAATAATAATAATA 8460  
QY 8461 CAGGTGTGTGAGATGACCAAGTATGCCAGTCTTCAAGAGGCTGAGGTGGAGATTC 8520  
DB 8461 CAGGTGTGTGAGATGACCAAGTATGCCAGTCTTCAAGAGGCTGAGGTGGAGATTC 8520  
QY 8521 CTGAGTGTGGAGGAGAGATTTGAGTGAACCTGATCTGTGCTTGCATTCACAGCTG 8580  
DB 8521 CTGAGTGTGGAGGAGAGATTTGAGTGAACCTGATCTGTGCTTGCATTCACAGCTG 8580  
QY 8581 GGTGACAGATTGAGACCTGTCTCAAAAACAAATTAATAATAATAATAATAATAATA 8640  
DB 8581 GGTGACAGATTGAGACCTGTCTCAAAAACAAATTAATAATAATAATAATAATAATA 8640  
QY 8641 CTGAGCAGGTAATTTCAAGTGGAAAACCTCCAGGGGAGTGAATGTCACTCACGCT 8700  
DB 8641 CTGAGCAGGTAATTTCAAGTGGAAAACCTCCAGGGGAGTGAATGTCACTCACGCT 8700  
QY 8701 GTATATCTGATACCGGCTAATTAAGAACTGTGTATACAGCAAGCACTAGGTAAT 8760  
DB 8701 GTATATCTGATACCGGCTAATTAAGAACTGTGTATACAGCAAGCACTAGGTAAT 8760  
QY 8761 ACTCAACAAATTTTGTGAGCATGTATGAAGTGGGCAATGCTCTAGGCACTGAGAT 8820  
DB 8761 ACTCAACAAATTTTGTGAGCATGTATGAAGTGGGCAATGCTCTAGGCACTGAGAT 8820  
QY 8821 ACAGTAGTCAATGACAGCAAGATGCTGCTGACAGGCTCTGCTAATAGTGAAGAG 8880  
DB 8821 ACAGTAGTCAATGACAGCAAGATGCTGCTGACAGGCTCTGCTAATAGTGAAGAG 8880  
QY 8881 GACATTAAGAAAGAAAGAAAGAAAGAAATTAATTTTGTATTTTGTATTTTGTATTTTGT 8940  
DB 8881 GACATTAAGAAAGAAAGAAAGAAAGAAATTAATTTTGTATTTTGTATTTTGTATTTTGT 8940  
QY 8941 GAAATTAAGACAGATAGTGGATAGAGGTGAGAAATGAGGGCTGTCTTCAAGAAA 9000

8941 GAAATTAAGA CAGGATAGTGGGATAGAGTGAAGAGTAAGGCTGCTTCTGAAGAAA 9000  
9001 TGAATTTTGAAGTGAAGCTTCAATGATGAGAAGGATTAACCAACAGATGCTGAGGA 9060  
9001 TGAATTTTGAAGTGAAGCTTCAATGATGAGAAGGATTAACCAACAGATGCTGAGGA 9060  
9061 AAGCAATTTTGAAGGAGGTGAGCAGCACAATATCTTCAAGAAATCAAGAAAGGCTGCTG 9120  
9061 AAGCAATTTTGAAGGAGGTGAGCAGCACAATATCTTCAAGAAATCAAGAAAGGCTGCTG 9120  
9121 AGCTGGAACA CAGAGAAAGAGAGCTGGGCTGACTTGAAGAGGAGGAGGAGCTGCTG 9180  
9121 AGCTGGAACA CAGAGAAAGAGAGCTGGGCTGACTTGAAGAGGAGGAGGAGCTGCTG 9180  
9181 AGCTGGAACA CAGAGAAAGAGAGCTGGGCTGACTTGAAGAGGAGGAGGAGCTGCTG 9240  
9181 AGCTGGAACA CAGAGAAAGAGAGCTGGGCTGACTTGAAGAGGAGGAGGAGCTGCTG 9240  
9241 GAGGAGCTGATATATTTCTAATCTATTTTATTAAGATCACTGACTTTTTCAGA 9300  
9241 GAGGAGCTGATATATTTCTAATCTATTTTATTAAGATCACTGACTTTTTCAGA 9300  
9301 ACATTAATTTTAAAGTACAGAGCTTGAAGAGAAATCCAGCTAGCAATCCGTCAGTT 9360  
9301 ACATTAATTTTAAAGTACAGAGCTTGAAGAGAAATCCAGCTAGCAATCCGTCAGTT 9360  
9361 GTCCAATTTAGAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9420  
9361 GTCCAATTTAGAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9420  
9421 TCACCAATTTATTTTGAAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9480  
9421 TCACCAATTTATTTTGAAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9480  
9481 GGT 9540  
9481 GGT 9540  
9541 TGAAGTCAAGAGTTCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9600  
9541 TGAAGTCAAGAGTTCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9600  
9601 TACCAATTTATTTTGAAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9660  
9601 TACCAATTTATTTTGAAGTGAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 9660  
9661 GGT 9720  
9661 GGT 9720  
9721 GCACTCAGCTGAGGAGCAAGAGTGAAGCTGCTGCTTCAATTAATTAATTAATTAATTA 9780  
9721 GCACTCAGCTGAGGAGCAAGAGTGAAGCTGCTGCTTCAATTAATTAATTAATTAATTA 9780  
9781 ATATCAGTCAAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9840  
9781 ATATCAGTCAAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9840  
9841 CAGCAATCTAGT 9900  
9841 CAGCAATCTAGT 9900  
9901 TGT 9960  
9901 TGT 9960  
9961 CATTAACATGTCAAGTGAAGGAGCTGATTTGAGAGCTTAAC 10000  
9961 CATTAACATGTCAAGTGAAGGAGCTGATTTGAGAGCTTAAC 10000

Sequence 917, Application US/10311455  
GENERAL INFORMATION:  
Publication No. US20030143606A1  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Level of Expression of Specific Genes  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 917  
LENGTH: 5234  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-917

Query Match 15.1%; Score 1509.4; DB 13; Length 5234;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 1777; Conservative 0; Mismatches 446; Indels 0; Gaps 0;

12 AGGCAAGGCCAGCAAAATCTGATGATTCAGACAGGCTTCTGACACTGAGCAAGATTG 71  
3012 AGGTAGGTTTATTAATTTTGAAGTATTAATAGGCTTTGATGATGAGTAAAGTTG 3071  
72 TTGAGGAATTTCTGATGAGATCATGGGCTGCTGAGAGGAGAAATTAAGTTTCA 131  
3072 TTGAGGAATTTTCTGATGAGATCATGGGCTGCTGAGAGGAGAAATTAAGTTTCA 3131  
132 AGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 191  
3132 AGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3191  
192 CACCCAGAAAGAGAGATGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 251  
3192 TATTTAGAAAGAGATGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3251  
252 GCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 311  
3252 GTAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3311  
312 TATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 371  
3312 TATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3371  
372 GCTGATGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431  
3372 GTTGTGATGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3431  
432 GTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 491  
3432 GTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3491  
492 GGGATGCTGAGAGGCTGAGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551  
3492 GGGATGCTGAGAGGCTGAGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3551  
552 GAGAGGAGGCTGAGAGCTGAGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 611  
3552 GAGAGGAGGCTTGTGAATTTTATTTTGTATGATGATGATGATGATGATGATGATGAT 3611  
612 TGAATCACTAGGCAAAATCTCAACAGAAATGGGAGCTTTTGGAGTGGGAGGAGGAGGAG 671  
3612 TGAATTTATAGGTAATTTTATTAAGAAATGGGAGTGGTGTGGAGTGGGAGTGGGAGGAG 3671

QY	672	ACTGAGGATACAGAGTGAATCCGACGAGAAAGATGGAATGTGGGGCCGAGACACTGGAGAG	731
Db	3672	ATTGAGGATGATGAGTGAATTTTAGGAGAGATGGAATGTGGGGTCCGAGATATTGGAGAG	3731
QY	732	AGAAACAGAGACTGTCAATTAAGGGCGTCTGTGACTCCAGATCTCATATTGCTACTAC	791
Db	3732	AGAAATAGGATGTGTAGATTAAGGGCGTGTGTGATTTTAAATTTTATTAATTAATTAAT	3791
QY	792	CATAACCTACCCCAATTCCTTAATTTCTCTACCTAGAGGGGGGAAATTGTCAGAA	851
Db	3792	TATAATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	3851
QY	852	TTTGGCTGCACACTAGCAACACTACTCAGTACTGGAATCATTTTGTGCAATTTTTTCA	911
Db	3852	TTTGGTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTTTGTATTTTTTTTA	3911
QY	912	TTCAACAATAATTTCTGGACAACCTTTATATGCCAGGCACTATTTTAGAGTACGGAT	971
Db	3912	TTTATATAAATTTTGGAAATTAATTTTATATGTAGGATTAATTTTAGAGTTAGGGAT	3971
QY	972	ATTATATGGTAAACAAGCAGGCAAAACAAAGCAAGCAACAACACATACACAGATTA	1031
Db	3972	ATTATATGGTAAATTAAGTAGGTAAATTAAGTAAATTAATTAATTAATTAATTAATTA	4031
QY	1032	GTAGACAGATGAAGAATTTCAAGTTTGTAGTAGTAAATTAACACAGAGGCTGAA	1091
Db	4032	GTAGATGATGAAGAATTTTAAGTTTATAGTAAATTAATTAATTAATTAATTAATTAAT	4091
QY	1092	ATGGCTAGATTAAGGCGGTCAAGAAAGCTTCATTGAGAGGTGACTTTAAGCAGAGTC	1151
Db	4092	ATGGTGTAGTAAAGGTGTAAAGAAAGTTTATTAAGAAAGGTAGTAAATTAAGTGAAGTT	4151
QY	1152	AGCTAGAAATATTGTGAATTCAGATTACAGTTCTAATTTGTTGGGTGGTTAAATTA	1211
Db	4152	AGTTAGAAATATTGTGAATTTTATAGTTATAGTTTATTTGTTGGGTGGTTAAATTA	4211
QY	1212	GCTTTTCCCCCAAGTGAACCTACAAAGAAAGCTAATTAATTAATTAATTAATTAATTA	1271
Db	4212	GTTTTTTTTTTTAAAGTGAATTAATTAAGAAAGCTAATTAATTAATTAATTAATTAAT	4271
QY	1272	TCTGGAAGAGAGACCCCTGTTTCTGCGCTCATTAAGTCAACCCCTCACITCOAGGCA	1331
Db	4272	TTTGGAAAGAGATATTTTGTGTTTGTATTAATTAATTAATTAATTAATTAATTAATTA	4331
QY	1332	CTTTTTCGAAAGCCCTTTCAGAGTCAGGAAAGCGCAGGCGATGCGGCTTGGACA	1391
Db	4332	TTTTTTTGAAGTTTGTGTTAGTTAGGAAAGCGAGAGTTGGGTATGGGGTTGGATA	4391
QY	1392	TTTGAACAAGTGAACATTAATGTCCCAAGCTCACTAGCCCAAGGTTAAAGCTGAAGA	1451
Db	4392	TTTGAATATATGAGATATTAATTTGTTTTTAAATTAATTAATTAATTAATTAATTAAT	4451
QY	1452	GGCTTTGGCATGCCCCAGAAAGGCCCTGATGAAGCTGGAATAAGCTGTCTCGAATA	1511
Db	4452	GCTTTGGGTAATTTTGAAGAAAGTTTTGATGAAGTTTGGAAAAAGTTGTTTTGAGTA	4511
QY	1512	TTTCTAAGTAAGTTATCTGTGTGTGTGTTACTAAAGTAGTAAGTATGCTGTCTTA	1571
Db	4512	TTTTTAAAGTAAGTTATTTGTGTGTGTGTTATTAAGTAGTAAGTATGTTGTGTTTA	4571
QY	1572	GCTGCTTAGAGCAGGCGCTGACACAGTACAGAGATTAATTAAGTTCCCTCCCTTTTCAC	1631
Db	4572	GTTGTTTTAGAGTACGGGTTTATTAAGTATTAAGTATTAATTAATTAATTAATTAAT	4631
QY	1632	CTCCCCATTGTGAGATTAATCTAATCACAAAAGTATCCTCAGTCTACTCTCC	1691
Db	4632	TTTTTTTATGTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4691
QY	1692	TGACTTAAGATGCTGTGAGACCATTTGCCAGTGTGAGATCAACGTGTGAGCTCAGACTG	1751
Db	4692	TGATTTAAGAGATTTGATTTATTTAGTGTGAGATTAATTAATTAATTAATTAATTAAT	4751

Oy	1752	TAGCCCACTAGCTGCTGAAAATATGCTGAAAGGGGGTTGGGGGGCAGCTGCCGGAAAAAG	1811
Db	4752	TAGTTTAGTTATTTGTTGAAAATTTGMAAGGGGTTGGGGGGTAGTTGTCCGAAAAAG	4811
Oy	1812	GAGCTTGGAATTCAGATTTCTGTCCAGACCCTGACCTATTTCAGGTGATGTAATCAACC	1871
Db	4812	GAGTTTTGGAATTTAGATTTTGTGTAATTTGATTTTAAATTTAGTAGCATGTAAATAGTT	4871
Oy	1872	AATATTGGCTTAGTCTCTGGAGACAGCAATTTCCAGTAGAGTTGAGAGTGGGGTGCTG	1931
Db	4872	AATATTGGTTAGTTTGGGAGATAGTAATTTTATAGAGTTGGAGAGTGGGGGTGCTG	4931
Oy	1932	CTGCTGCCAATCTATATNAGGAGATTTCAATGTGCACCCAGAGCTGTCCTGGCTGTG	1991
Db	4932	TTGTTGTTAATTTTATATNAGGAGATTTTAAATTTGATTTTAAAGTTGTTTGTGTTTGT	4991
Oy	1992	CAGCTCAGACATGGCTAGAGGTACTGGAGACCCGTTGCACTGGGGTTGTGGAGCTATGC	2051
Db	4992	TAGTTTAGATATGGTTAGGGTAATGGAGATTTGCTGTATTTGGGTTGTGGAGTTTATGT	5051
Oy	2052	TGCTCTCGGCATTGCGACCCCTCTTCTCCGTGAGTAAAGCTGGAGCTGAAGCGAAG	2111
Db	5052	TGGTTTTTGATATGTGTAATTTTTTTTTTTTCGTGAGTAAAGTTGGGATTGAAGCGAAG	5111
Oy	2112	GATTGAGTTCTGGGCTNAGGGTAAGGTNAGGCGCAGTTTTCAGGCTCGGTCAAAATTTGGGG	2171
Db	5112	GATTGAGTTTGGGGTTAGGGTAAGGTNAGGGTTAGTTTTCAGTTTCGGTTAAATTTGGGGG	5171
Oy	2172	TCAGGGGCTATGGGAAAAGGATCGGTCCCAATGGAATCAAGATATCTATTTTGTCTCCCT	2231
Db	5172	TTTAGGGGTATNAGGAAAAGGATCGGTTTAAATGGAATTAAGATATTTATTTTGTGTTTTTTT	5231
Oy	2232	AGG 2234	
Db	5232	AGG 5234	

```

RESULT 3
US-10-311-455-918/c
; Sequence 918, Application US/10311455
; Publication No. US2003014366A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PISENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 918
; LENGTH: 5234
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-918

```

Query Match	12.8%	Score 1281.6	DB 13	Length 5234
Best Local Similarity	73.4%	Pred. No. 0		
Matches 1638	Conservative	0	Mismatches 554	Indels 0
			Gaps 0	

  

OY	1	TCCTCTTCCCGAGCGCCGACAAATCTGTGAGATTCAGACAGGGTTCTGACAGCTG	60
DB	2234	TCCTCTTCCCGAAACCAACCCACAAATCTATATAATTCAAACAAATTTAACTAACTA	2175



Db 14 TTATCTCCCTA 3

RESULT 4  
US-09-918-995-31396  
; Sequence 31396, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31396  
; LENGTH: 533  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(533)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31396

Query Match 4.6%; Score 457.6; DB 11; Length 533;  
Best Local Similarity 98.9%; Pred. No. 7.1e-108;  
Matches 460; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4479 GGAAGAAGATACCAAGTCTCCCAAGATATTTCTGGATCCCATCCCACTGGAT 4538  
DB 68 GGAAGAAGATACCAAGTCTCCCAAGATATTTCTGGATCCCATCCCACTGGAT 127  
QY 4539 GCAAGTGTGAATGTCAAGTGTGAAGATGTCAAGTGTGAAGTGTGAAGTGTGAAGT 4598  
DB 128 GCAAGTGTGAATGTCAAGTGTGAAGATGTCAAGTGTGAAGTGTGAAGTGTGAAGT 187  
QY 4599 CAGTCCAGCTGTGAATGTCAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGT 4658  
DB 188 CAGTCCAGCTGTGAATGTCAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGT 247  
QY 4659 GATGTGTGAATGT 4718  
DB 248 GATGTGTGAATGT 307  
QY 4719 AATATAGAAGCAAGT 4778  
DB 308 AATATAGAAGCAAGT 367  
QY 4779 GTGAAGCAAGT 4838  
DB 368 GTGAAGCAAGT 427  
QY 4839 GGAAGCAAGT 4898  
DB 428 GGAAGCAAGT 487  
QY 4899 CCCTTCAAGT 4943  
DB 488 CCCTTCAAGT 532

RESULT 5  
US-09-918-995-32463  
; Sequence 32463, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32463  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(495)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-32463

Query Match 4.5%; Score 445.2; DB 11; Length 495;  
Best Local Similarity 99.3%; Pred. No. 1.1e-104;  
Matches 447; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4908 TTGACTTTTATATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 4967  
DB 46 TGTGCTTTTATATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 105  
QY 4968 GTGACTGATTTGTGCTGATGCTTTTGTGATCTTTTGTGCTGATGCTTTTGTGCTGATGCT 5027  
DB 106 GTGACTGATTTGTGCTGATGCTTTTGTGATCTTTTGTGCTGATGCTTTTGTGCTGATGCT 165  
QY 5028 TGTGCTTTTATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 5087  
DB 166 TGTGCTTTTATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 225  
QY 5088 TTCCAGACTTCTCCAGCTGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 5147  
DB 226 TTCCAGACTTCTCCAGCTGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 285  
QY 5148 TGTGCTTTTATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 5207  
DB 286 TGTGCTTTTATATCTTTTACCTTAACATGATCCAGAAACCTATGCTATCTCT 345  
QY 5208 TACCATTTATCTGCTGATGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 5267  
DB 346 TACCATTTATCTGCTGATGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 405  
QY 5268 TCACCAATGCTGCTGATGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 5327  
DB 406 TCACCAATGCTGCTGATGCTTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 465  
QY 5328 TACGGGAACATGAAG 5387  
DB 466 TACGGGAACATGAAG 495

RESULT 6  
US-09-918-995-31405  
; Sequence 31405, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31405  
; LENGTH: 480  
; TYPE: DNA  
; ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(480)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31405

Query Match 4.4%; Score 441.6; DB 11; Length 480;  
Best Local Similarity 99.1%; Pred. No. 9.5e-104;  
Matches 444; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4908 TTGACCTTTATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 4967  
DB 32 TCTGACTTTATATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 91  
QY 4968 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTGCGGTTCCATCTCTCT 5027  
DB 92 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTCTGCGGTTCCATCTCTCT 151  
QY 5028 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 5087  
DB 152 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 211  
QY 5088 TTCCAGACTCTCCCACTGCTTTGCTGAGCTGATCTGCTGTGTTTGTCTGGAATT 5147  
DB 212 TTCCAGACTCTCCCACTGCTTTGCTGAGCTGATCTGCTGTGTTTGTCTGGAATT 271  
QY 5148 TGTGTATGATCAACCTCTTGAACCATCTCTGATGCTGAGCTCATGTGCTCAG 5207  
DB 272 TGTGTATGATCAACCTCTTGAACCATCTCTGATGCTGAGCTCATGTGCTCAG 331  
QY 5208 TACCATTTATCTGGCCCATCTCTGGGACCCAGAGAAAGACAAAGAGGCGTTAACCCGGTTC 5267  
DB 332 TACCATTTATCTGGCCCATCTCTGGGACCCAGAGAAAGACAAAGAGGCGTTAACCCGGTTC 391  
QY 5268 TCACCAAAATGCTGTGATTTGATTTGACCAAGGAGACCGGAGTGTCTGGGACTTGGC 5327  
DB 392 TCACCAAAATGCTGTGATTTGATTTGACCAAGGAGACCGGAGTGTCTGGGACTTGGC 451  
QY 5328 TACGGGAACCATGAAGAGGCTTCTGG 5355  
DB 452 TACGGGAACCATGAAGAGGCTTCTGG 479

## RESULT 7

US-09-918-995-31407  
Sequence 31407, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31407  
LENGTH: 494  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(494)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31407

Query Match 4.4%; Score 441.4; DB 11; Length 494;

Best Local Similarity 99.8%; Pred. No. 1.1e-103;  
Matches 442; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4908 TTGACTTTTATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 4967

DB 52 TCTGACTTTTATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 111  
QY 4968 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTGCGGTTCCATCTCTCT 5027  
DB 112 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTCTGCGGTTCCATCTCTCT 171  
QY 5028 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 5087  
DB 172 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 231  
QY 5088 TTCCAGACTCTCCCACTGCTTTGCTGAGCTGATCTGCTGTGTTTGTCTGGAATT 5147  
DB 232 TTCCAGACTCTCCCACTGCTTTGCTGAGCTGATCTGCTGTGTTTGTCTGGAATT 291  
QY 5148 TGTGTATGATCAACCTCTTGAACCATCTCTGATGCTGAGCTCATGTGCTCAG 5207  
DB 292 TGTGTATGATCAACCTCTTGAACCATCTCTGATGCTGAGCTCATGTGCTCAG 351  
QY 5208 TACCATTTATCTGGCCCATCTCTGGGACCCAGAGAAAGACAAAGAGGCGTTAACCCGGTTC 5267  
DB 352 TACCATTTATCTGGCCCATCTCTGGGACCCAGAGAAAGACAAAGAGGCGTTAACCCGGTTC 411  
QY 5268 TCACCAAAATGCTGTGATTTGATTTGACCAAGGAGACCGGAGTGTCTGGGACTTGGC 5327  
DB 412 TCACCAAAATGCTGTGATTTGATTTGACCAAGGAGACCGGAGTGTCTGGGACTTGGC 471  
QY 5328 TACGGGAACCATGAAGAGGCTT 5350  
DB 472 TACGGGAACCATGAAGAGGCTT 494

## RESULT 8

US-09-918-995-31269  
Sequence 31269, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31269  
LENGTH: 478  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(478)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31269

Query Match 4.4%; Score 440.8; DB 11; Length 478;  
Best Local Similarity 99.5%; Pred. No. 1.5e-103;  
Matches 442; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4908 TTGACTTTTATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 4967  
DB 34 TCTGACTTTTATATACCTTTTACCTTAACAAGCTCCAGAAACCTATGCTATCTCT 93  
QY 4968 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTCTGCGGTTCCATCTCTCT 5027  
DB 94 GTACACTGAGTTTGTCTGCATGCTTTGGAACTTTCTCTCTCTCTGCGGTTCCATCTCTCT 153  
QY 5028 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 5087  
DB 154 TGTGTGCTTTTAAATTCCTAATTCATGATTCATCTTAAGTATCATCTTCCCTGGAAATT 213





Db	11815	-----ACAGTGGCTCAGCGCCCGTAATCCAGCATTTTGGGAGGCTCAAGCAGGTGATAA	11869
QY	8385	CTTGAGTCCAGGAGTTCTAGACAGCCTGGGCAACATGTGGAAATCTCATCTCTATAA	8444
Db	11870	TTTGAGGTGCGAAGTTTGAGACCAAGCCTGGCCAACTGGGGAAACCTGTCTCTATAA	11929
QY	8445	AATACAAAAATTAGCCAGGTGTGTGGCAATGCACAGTAGTCCCGAGCTGTTCAGAGGC	8504
Db	11930	AATACAAAAATTAGTCTAGGACGATGATGTGTGCACACCTGTATCCGAGCTAATCTTGGGGAC	11989
QY	8505	TGAGTGTGGAGGATTTGCTCGAGTGTGGGAGGACAGATTGCAGTGAACCGTGACTGTGC	8564
Db	11990	TGAAGCAGGAGAACTTTGAACTGTGAGGCGGAGGTGTGTGTGAGCCGAGATCATACC	12049
QY	8565	TCTGCAATCCAGCCTGGGTGACAGATTGAGACCCCTGTCTCAAAAAACAATTAATTAAT	8624
Db	12050	ACTGCATCTCAGTCTGGGTGACAGAGTGAAGACCTTATCCCAAAAAATTAAAAA	12109
QY	8625	AAA 8627	
Db	12110	AAA 12112	

```

RESULT 10
US-10-017-161-1795/c
Sequence: 1795, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIHO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABRATANI, HITROYUKI
TITLE OR INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 08435/0152
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1795
LENGTH: 43419
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: source
LOCATION: (1)..(43419)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(278)
FEATURE:
NAME/KEY: CDS
LOCATION: (8419)..(8609)
FEATURE:
NAME/KEY: CDS
LOCATION: (110482)..(10594)
FEATURE:
NAME/KEY: CDS
LOCATION: (19132)..(19408)
FEATURE:
NAME/KEY: CDS
LOCATION: (41900)..(42106)
FEATURE:
NAME/KEY: CDS
LOCATION: (42300)..(42393)
FEATURE:
NAME/KEY: CDS
LOCATION: (43094)..(43219)
FEATURE:
NAME/KEY: modified_base
LOCATION: (20434)..(20533)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1795

```

Query Match	3.7%;	Score 372;	DB 13;	Length 43419;
Best Local Similarity	56.6%;	Pred. No. 2.2e-84;		
Matches 852;	Conservative	0;	Mismatches 535;	Indels 119;
				Gaps 10;

[illegible]



```

QY 8157 TTTCCAGAGAGGGAATCTGAATGCCCCAGTCCATATTTTCAGACACACCATTTAAAGT 8216
DB 20404 CTCTCTGTGGCCAGGCTGGAATGCAATGATCTTGCTCACTGCACTCTGCT 20345
QY 8217 GGTGATGCGACGCTATGATGCTATCACTTAATGGGTTGGAACTCATCTTACTTC 8276
DB 20344 CCGGGTTTCAAGGATCTCTGCTCCAGCTCCCAAGTAGCTGGGACTCAAGGATGT 20285
QY 8277 ATTGCACAAAGACGATAGCTGCTGTTCTC-AAATATAGGCCCCCTGGGCAAGTGTGTG 8335
DB 20284 GCCACTGACCCGGCCAAATTTTAAATTTTAAATCAAAAGTTGGGCGCGGACGTTG 20225
QY 8336 GCTCATGCTATATATCCCAACACTGTGGAGCCGAGGGGGGAGAT----- 8382
DB 20224 GCTCAACCTGTATCTCCAGCACTTTGGAGGCCAAAGTGGGGGATCAACCACTGACTT 20165
QY 8383 -CACTTGAAGTCCAGAGTTCTTGAAGCAAGCTGGGCAACATGTGAATCTCATCTTACT 8441
DB 20164 CCACTGGAAGTCAAGACTTTTGAAGCAAGCTGGGCAACATGTGAATCTCATCTTACT 20105
QY 8442 AAAAAATACAAAAAATATAGCAAGTGTGTGGGATGACCACTGATCTCCAGCTTTACGA 8501
DB 20104 AAAAAATAC-AAAAATAGCCAAAGTGTGTGGGCGCGCCCTGTAGCCCAAGTACTCAGGA 20046
QY 8502 GCGTGAAGTGGAGATGCTGCTGAGTGTGGAGGAGAGATTCAGTGAACGCTGACTGT 8561
DB 20045 GCGTGAAGTGGAGATGCTGCTGAGTGTGGAGGAGAGATTCAGTGAACGCTGACTGT 19986
QY 8562 GCGTCTGCAATCCAGCTGGGTGAGCAAGTGTGAAGCCCTGTCTCAAAAAAATAATATA 8621
DB 19985 GCGACTGCACTCCAGCTGGGTGAGCAAGTGTGAAGCCCTGTCTCAAAAAAATAATATA 19926
QY 8622 AATATA 8627
DB 19925 TACAAA 19920

```

## RESULT 11

US-10-017-161-1795

```

/ Sequence 1795, Application US/10017161
/ Publication No. US20030143668A1
/ GENERAL INFORMATION:
/ APPLICANT: SUMA, MAKIKO
/ APPLICANT: ASAI, KIYOSHI
/ APPLICANT: AKIYAMA, YUTAKA
/ APPLICANT: ABURATANI, HIROYUKI
/ TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
/ FILE REFERENCE: 084335/0152
/ CURRENT APPLICATION NUMBER: US/10/017,161
/ PRIOR FILING DATE: 2002-12-18
/ PRIOR APPLICATION NUMBER: JP 2001/246789
/ NUMBER OF SEQ ID NOS: 2430
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 1795
/ LENGTH: 43419
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: source
/ LOCATION: (1)..(43419)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (201)..(278)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (8419)..(8609)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (10482)..(10594)
/ NAME/KEY: CDS
/ LOCATION: (19132)..(19408)

```

```

/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (41900)..(42106)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (42300)..(42393)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (43094)..(43219)
/ FEATURE:
/ NAME/KEY: modified base
/ LOCATION: (20434)..(20533)
/ OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1795

```

Query Match 3.7%; Score 368.2; DB 13; Length 43419;  
 Best Local Similarity 59.5%; Pred. No. 2,2e-83;  
 Matches 866; Conservative 0; Mismatches 48; Indels 111; Gaps 10;

```

QY 7269 CTTGAGCCTGAGAGTTTGAAGACAGCTGGGCAACATGCAAAACCTCATCTTACAAA 7328
DB 12442 CTTGAAGCCAGAGATTGAGACAGCTGGCCAACTGGTGAACCCATCTTACAAA 12501
QY 7329 AATATATAAATTAGTGGGTGTAGTGCCTTCTATAGTCCATCTTACTTCAAGGCT 7388
DB 12502 AATACAAAATTAGCAGGCAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12561
QY 7389 GAG-----CCGGAAGTGAAGCTTCAAGTGAACCTGATCTGTCTA 7430
DB 12562 GAGTCAGAGAAATCCCTTGAACAGGAGAGAGAGTTTCAAGTGAATGATCCGACCA 12621
QY 7431 CTGCACTCCAGCTGGGTGAGAGTGAACCATGCTCAAAA-AAAAAACAACAAA 7489
DB 12622 CTGCACTCCAGCTGGGTGAGAGTGAACCATGCTCTGTCAAAAACAACAAA 12681
QY 7490 AAAAAACAACAAAACAACAAAACAACAAAACAACAAAACAACAAAACAACAAA 7529
DB 12682 ACAACCTGAGCTGGGTATGATTAAGAAAGAGTTTAAATGCTCAAGTTCTGAG 12741
QY 7530 --TATATATATATCTAGTGAAGTGAATGCACTATTTTGTGAATCAACCAATGAC 7587
DB 12742 GCTGTAAAGCATGACACAGCATCTGCTGGCTTCTGGGAGGCTTCAAGAGCTTTTA 12801
QY 7588 CCAAGTACAGATGGGCGAGTCCCTCCCTCCTCAGTGAATTTT-----CT 7636
DB 12802 CTTTGGCAAAAGGGTGAAGGAGCAACACATCTAGACAGAGCAAAAGAGAGAG 12861
QY 7637 TTCTGACTCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7696
DB 12862 TGTGGGAGGTGACATATCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12921
QY 7697 TCACCAAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 7756
DB 12922 TCACCAAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 12981
QY 7757 TCAGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7816
DB 12982 TCAGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13041
QY 7817 GCTGGCTAATTTTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7876
DB 13042 GCGCGCTAAT-----GTTT 13057
QY 7877 TGTATTTTATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 7936
DB 13058 TGCATTTTATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 13117
QY 7937 CTCAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7996
DB 13118 CTC--GTATTCGCGCACTGAGCTGCAAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13175
QY 7997 CGCTGGCCCTCATAT--GGTTTATCTATCTCTTGTCTTCAACATTTGGCTT 8053

```

Db	13176	CGCCAGCAGATGCGACATACCTTTAAACAATAGATCTGGAAAGAACTCATATCTCGA	13233
Qy	8054	GCACGTGACCAT-----CATGTTCTCTCCACTTTCTCACTACTTCATGATCTTTTCAATC	8108
Db	13336	GAACAGCAACCAATCCATGACGAGATCCCATCCCATGACCCCAAAACCTCCACCAATGCGCA	13295
Qy	8109	TCAGTTCCAACTGATAC-----CTCCCTCAGTGCCTTTTTTCTCTGTAAGATTTTCCAGA	8164
Db	13296	CCTCCAAACCTGGGGAATTAACAATTCACATGAGATTTGGGCAAGGACAGATTAAGCAAT	13355
Qy	8165	GAGGGAATCTGAATGGCCCAATCCATATTTTTCAGACCAACCAATTAAGGTGTGATT	8224
Db	13356	GATAGGTATCGTCCACATATACAGGTATGAGGGCTACCTGAGGAGGCAACAGCTGAAGTGA	13415
Qy	8225	GCCAGCCTATGATATGGCTACATTAATGGGTGGGAACTCATATTTACTTCATTCACA	8284
Db	13416	AGGCTTGTCTGTGAAGCTATGCGGGCTTTTATGAGGAAGGCTCTATGCAAGGTGAGGA	13475
Qy	8285	AAGCAGCATAGCTGTGCTTCACAAATAGGGGCCCTGGGCGCAGATGATGCTCATGCGC	8344
Db	13476	TGGAATATGGGGCTTGATGCTTAGAAAGTGGAGAGAAAGGCAGATGCAATGCTCACACC	13535
Qy	8345	TATATATCCCAACACTGTGGAGGCGCGAGGGGCGCAGATCACTTGAATCCAGAGTTCAG	8404
Db	13536	TGTATTTCCCAAGCACTTTTGGGAGGCGCGAGGTGGGTGATCATCTGAAGTCAGAGTTCCAG	13595
Qy	8405	ACCAAGCTGTGGCAACATGTGAAATCTCATCTTCTATAAAATATACAAAAATTTAGCCAG	8464
Db	13596	ACCAAGCTGTGGCTAATGATGTGAACCCCATCTCTATTAATAATA-AAAAATTTAGCCAG	13654
Qy	8465	TGTGTGTGCATGACCAAGTACTCCCAAGCTGTTCAAGAGGTGAGGTGGAGGATTTGCTCG	8524
Db	13655	CATGTGTGCGTACCACTGTATCCCACTATCAAGAGGCGCAGGAGCAAGAAATTTGCTTG	13714
Qy	8525	AGTGTGGAGGACAGAGATTCAGATGAACCGTGACTGTGCTCTGCATCAAGCTGGGTG	8584
Db	13715	TACCTGGAGGTAGATTTTCAATGAGACCAAGATTTGTGCATTGCACTCAAGCTGGGGG	13774
Qy	8585	ACAGATTGAGACCTGTCTCAAAAAACAATAATATTAATAATTAATATATGTGTTCTGA	8644
Db	13775	ACAAGAGAGAGCTCCGTCTCAAAAAAATAAAAAAAAAAAAAAGAAAGTGGAGAGAAAG	13834
Qy	8645	GCAGGTTAATTTTCAAG 8659	
Db	13835	GAACGTGACTGACG 13849	
RESULT 12			
US-10-374-979-11			
Sequence 11, Application US/10374979			
Publication No. US20030219793A1			
GENERAL INFORMATION:			
APPLICANT: John P. Carulli et al.			
TITLE OF INVENTION: THE HIGH BONE MASS GENE OF 11q13.3			
FILE REFERENCE: 032796-021			
CURRENT APPLICATION NUMBER: US/10/374,979			
PRIOR FILING DATE: 2003-03-04			
CURRENT APPLICATION NUMBER: US 09/544,398			
PRIOR FILING DATE: 2000-04-05			
PRIOR APPLICATION NUMBER: US 09/543,771			
PRIOR FILING DATE: 2000-04-05			
PRIOR APPLICATION NUMBER: US 09/229,319			
PRIOR FILING DATE: 1999-01-13			
PRIOR APPLICATION NUMBER: US 60/071,449			
PRIOR FILING DATE: 1998-01-13			
PRIOR APPLICATION NUMBER: US 60/105,511			
PRIOR FILING DATE: 1998-10-23			
NUMBER OF SEQ ID NOS: 109			
SEQ ID NO 11			
LENGTH: 66933			
TYPE: DNA			
ORGANISM: Homo sapiens			
US-10-374-979-11			

Query	Match	Local Similarity	3.5%: Score 35.5; DB 13; Length 66933;
Best Local Similarity	59.6%;	Prod. No. 7, 8e 80;	
Matches	871;	Conervative	0; Mismatches 460; Indels 130; Gaps 11;
Qy	7184	TTAAAAATACCAAGCCGGGCGTGTGAGTCAAGCCCTGTAATCTAGACTTTGGAGGC	7243
Db	61260	TTAAGAACTTTGGGGCCCAAGTGAATGGCTTGGACGTGTATCTCAGCACTTATAGAGGC	61311
Qy	7244	TGAGGCAAGTGAGATGTCTTAGGGGCTTAGCCCTAGAGTTTAGACCAAGCCCTGGCAAC	7303
Db	61320	TGAATGAGGGATCACTTAGG-----CCGGAGTTTAGACCAAGCCCTGGCCAC	61377
Qy	7304	ATGCAAAACCTCATCTCTCAAAAAATATAAAAATTAGTCGGGTGTAGTACGCTCC	7363
Db	61371	ATGGAAGAACCCGGCTCTAGTAAAAATTAAAAATTAGCCGGGTATGGATCCAGCT	61433
Qy	7364	TATATG-----TCCATCTACTCAGAGGCTGAGCCCGGAAGTTCAGAGCTTCAGTAGCC	7418
Db	61431	ACTTGGGAGTCTGAGAGATGAGATTTGCTTGAACATGGAGGGGAGGTTCAGTAGGCC	61490
Qy	7419	GTAATCGTCTACTGCACTCCAGGCTGGGTGACAGAGTAGACATAGTCTCAAAAAAAG	7478
Db	61491	GAGATCCGCGATTGCACTCCAGCTGGGCGACAGAGCAGACTCTGTCTCAAAAAAAG	61555
Qy	7479	AAAAACAAAAACAAAACAAAACAAACAAACAAAAAACCATATATATATATAT	7538
Db	61551	AAAAAAGAAAAAAGAAAAAATAATATCATCTCTTTTATGAGCATATATATATAT	61610
Qy	7539	ACCTAGCTGAGGTGAGAAATGACATATTTGTGTAAATATCAACAATAGCCAGCTACAG	7598
Db	61611	AT	61667
Qy	7599	ATGGGAGAGTCCCTCCCTCTCACTGGTAATTTTCTTCTCTGACTCAGATTGTTGT	7658
Db	61664	---AGGCTTCCCGTCTGATCATATAAAAAACAATTATTTTCACTCTCTCTTTT	61722
Qy	7659	GTTGT	7717
Db	61721	TT-----TTTGAAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	61766
Qy	7718	GGCGCAATCTTGGTCACTGCAACTCTGCTCTGGGTCAAGCATCCTCTGCTCA	7777
Db	61767	GGCGCAATCTCAGCTCACTGTAACCTCCGCTCCGGGTGAGATATCTCTGCTCA	61822
Qy	7778	GCTCCCGTATATGCTGGGACTACAGGGGCAATACCAATGCTGCTAATTTTGTATTT	7837
Db	61827	CTTCCCGGATATGCTGGGATTTATAGGCAATGCAACCAATGCTGGCTAA-----	61875
Qy	7838	TTTTTTGGGTACATATATATTTTATTTTATTTTGTATTTTGTATGAGAG-TAG	7896
Db	61876	-----TTTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	61899
Qy	7897	GGTTTCACATTTTGGGCAAGGCTGTCTCGAATCTCTGACTCAGGTGATTCGGCTGCT	7956
Db	61900	GGTTTCCATATTTTGGGCAAGGCTGTCTCGAATCTCTGACTCAGGTGATTCAGCACT	61955
Qy	7957	CGGCTCTCCAAATGCTGAGATTTACAGGCAATGAGCAACCAAGCTGCTGCTCATAGTT	8016
Db	61960	CGGCTCTCCAAATGCTGAGATTTACAGGCAATGAGCAACCAAGCTGCTGCTCATAGTT	62019
Qy	8017	TTTTATCATTTCTGTGCTTCTTCAACAATTGAGCTTGAACGTGACATCATGTTCTCT	8076
Db	62020	TCTTATGATCT	62055
Qy	8077	CCACTTTCATCACTTCAATGATCTTTCAGTCTCAGTTCCAACTGATATCTCTCAGTT	8136
Db	62059	TAAATCTTCTGGGATGTAATTAATTTGATGAGCAAGTGGGATCACTGTTGTTCTTG	62118
Qy	8137	GCTCTTTTTCCTATGATTTTCCAGAGAGGAATCTGATAGGCCAGTATATTTTC	8196
Db	62119	GCTGATGCTTATGAGGTGGGCTGATATTTATGCTGGGGTCTATCAGAGAGCAAACTCTATG	62178

QY	8197	AGACCAACACACATTAAAGGTGGTTATGCAAGCTATGATTAATGGCTACATTAAAGGGTT	8256
Db	62179	AG-----AATTGAACAGAAAAGTTCCGTCTACAGGCTTATTACAGGACTGGAA	62230
QY	8257	GGGAATCATCATTTACTTACTTCATTGCAAAAGCAGCATAGCTCTGGTTCTCAAAATAGGCG	8316
Db	62231	TAGCAGAAATTTGAACAGTGAGATGTCAGAGAACTCT-----AAGAAATGCAAG	62277
QY	8317	CCCGGGCCAGAGTGGTGGCTCATGCTATATCCCAACATCTGGGGAGGCGCAGGGGG	8376
Db	62278	GAATAGGCCAGGCAATGGTGGCTCAACCTGTCAATCCAGCACTTTGGGAGACCAAGGCGG	62337
QY	8377	GCAGATCATCTTGAATCCAGAGTTCTAGACCAAGCCTGGGCAACATGTAATCTCATCT	8436
Db	62338	GTGGATCACTGAGGTCAGAGTGTCCGAGACCAAGCTGGCAACATATGTGAACCCCATCT	62397
QY	8437	CTTACTAAAAATAC-AAAAAATTAGCCAGGTGTGGTGGCATGACCAAGTATGTCCTCCAGCTGT	8495
Db	62398	CTTACTAAAAATTAACAAAAAATTATGCTGGGTGTGGTGGCCCATGCTCTGATATCCCAAGCTAC	62457
QY	8496	TCAAGAGGCTGAGGTGGAGGATGCTCGAGTGTGGAGGACAGAGATTGCAGTGAACCGT	8555
Db	62458	TCCGGAGGCTGAGGCGAGAGAAATCACTTGAACCTGGAGGCGAGAGGTTCAATGAGCCGA	62517
QY	8556	GACTGTGCTCTGCGAATCCAGCCTGGGTGAACAGATTGAAGCCCTGTCTCAAAAAACAAT	8615
Db	62518	GATCATGCTCACTGATCCAGGCTGGGTGGAGAGCGAACTGTCTGTAAAAAAATAAAA	62577
QY	8616	AAATTAATTAATTAATTAATG	8636
Db	62578	AAAAACAAGATTCAACTTG	62598

```

RESULT 13
US-09-764-891-8396
; Sequence 8396, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8396
; LENGTH: 27154
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8396

Query Match      3.5%; Score 347.2; DB 11; Length 27154;
Best Local Similarity 58.6%; Pred. No. 4.5e-78;
Matches 899; Conservative 0; Mismatches 483; Indels 152; Gaps 11;

OY      7195 CAGGCCGGGCGTGGTGAATC-ACGCGCTTAATCTCTAGACATTTGGAGAGCTGAGGCAAGT 7253
           |||||
Db       542 CAGGCCGGGCGTGGTGGCTCAACGCGCTTAATCCAGACATTTGGAGGCGCAAGGCAAGT 601

OY      7254 GGATTCGTTGAGCGGCGTTGAGCCCTTAGAGATTGAGACCAAGCTCGGGAACATGCGCAAAAC 7313
           |||||
Db       602 GGATTCACCTGAGATC-----AGGAGTTCAGAGACCAAGCTTAGCCAAACATGGTGTGAAC 652

OY      7314 CTCATCTCTCAAAATAATATAATAATTAGTCGGGAGTGTAGTCGTTCTATAGTCCCA 7373
           |||||
Db       653 C-CATCTCTAATAAATAATACAAATAATTAGCCAGGCGGTATATGGGCACTGTATATCCCA 711

OY      7374 TCTACTTCAGAGGCTAG-----CCCGAAGGTCAGAGCTTCAGTGA 7415
           |||||
Db       712 GCTACTCGGGAGGCGTGAAGGCAAGAAATCGCTTGAACCAAGGAGGCAAAAGTTGCAAGTA 771

OY      7416 GCCGTGATCGTGCTACTGCACTCCAGCGCTGGGTGACAGAGTGAGACCATGTCCTCAAA-- 7473

```

[illegible]

Db 1804 GGTGAAACCTATCTACTATAAAATAC-AAAAATTAGCCGGCTGTGGCAAGCACT 1862  
Qy 8482 GTATGCCAGCTGTTGAGAGGCTGAGGTGGAGATGCTCGAGTGTGGAGGCAAGA 8541  
Db 1863 GTATATCCAGCTACTTGGAGGCTGAGGCAAGAAATCACTTGAACCTCAGAGGCGGATG 1922  
Qy 8542 TTGAGTGAACCGTACTGTGCTCTGCAATCCAGCCTGGGTGACAGATTGAGACCTGT 8601  
Db 1923 TTGAGTGAAGCTGAGATCGAGCACTGCACTCCAGCCTGGGCAACAAAGGAGACTCAT 1982  
Qy 8602 CTCAAAAACAAATTAATTAATTAATAT 8635  
Db 1983 CTCAAAAACAAATTAATTAATTAATAT 2016

## RESULT 14

US-10-034-650-10/c  
; Sequence 10, Application US/10034650  
; Publication No. US20030216558A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 529452000128  
; CURRENT APPLICATION NUMBER: US/10/034,650  
; PRIOR FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 09/474,377  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 53106  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-034-650-10

Query Match 3.5%; Score 346; DB 13; Length 53106;

Best Local Similarity 58.9%; Pred. No. 1,4e-77;  
Matches 850; Conservative 0; Mismatches 500; Indels 94; Gaps 11;

Qy 7192 ACCGAGCCGGGCTGTGATGATCGCTGTAATCTAGACATTGGAGGCTGAGGAG 7251  
Db 32946 ACATCGGCTGGGACGGTGGCTATGCTTAATCCAGCACTTGGAGGCGCAAGGAG 32887  
Qy 7252 GTGGATTGCTTGAAGCGGCTTGAAGCTTGAAGATTCAGACCAAGCTGGGCAATGSCAAA 7311  
Db 32886 GTGGATCGCTGAGGTC-----AGGATTCAGACCAAGCTGACCAATATGATGAA 32836  
Qy 7312 ACCTCATCTCTCAAAAATAATAATAATAGTGGGCTGTGATGCTGCTTAATCTCC 7371  
Db 32835 ACCCGCTCTTCAAAAATAATAATAATAGTGGGCTGTGATGCTGCTTAATCTCC 32776  
Qy 7372 CATCTACTGAGAGCTGAGCCCGGA-----AGGTGAGGCTTCACT 7413  
Db 32775 CAGCTACTGAGAGGCTGAGAGCAAGATCACTTGAACATGGAGGCGAGAGTGTGAGT 32716  
Qy 7414 GAGCGGTGATGCTGCTACTGCACTCCAGCTGGGTGAC-AGAGTGAAGCAATGTTCAAA 7472  
Db 32715 GACCTGAGATTGTGCTCACTCCAGCTGGGCAACAGAGGAAATCTCACTTCAAA 32656  
Qy 7473 AAAAAACAAAACAA-----AAACAAACAAACAAACAAACAAACAAACCAATAT 7526  
Db 32655 AAAAAACAAAACAAAGCTTGGCATTTAAATCACTTGAATGAGATTAATGTCACAC 32596  
Qy 7527 ATATATATATATATCTAGCTGAGGTGAGATGCACTATTTTGTGAATCAACCAACATGA 7586  
Db 32595 GAGTTTGTGGGAGCACTTCAACTTTAGCACTAGATATTTCTGTTATGTTATTTTAG 32536  
Qy 7587 CCAAGCTACAGCATGGGAGAGTCCCTCCCTCACTGATTAATTTTCTTCTGACT 7646

Db 32535 CTTAATCTCTTCAATTTCTAACAATTAAGATCCAGATTAATCCATATATTTGGTTCT 32476  
Qy 7647 CACAGTTTGTGTTGTTGTTGCTGTTGTTGATGAGAGTCACTCTGTCACCCAGGC 7706  
Db 32475 TTCTTTTGTGTTTGTGTTTGTGTTTGTGAGACAA--GAGTCTGCTCTGTTC-GCCAGGC 32419  
Qy 7707 TGAAGTGAATGAGCGCAATCTTGGTTCACTGCAACCTCTGCTCTGAGTTCAAGCATC 7766  
Db 32418 TGAAGTGAATGAGCATGATCTGAGTCACTGTAACCTCTGCTCTGAGTTCAAGCATC 32359  
Qy 7767 CTCTGCTGAGCTCTCCGTTATGCTGAGACTACAGGCGCATACCAATGCTGGCTAA 7826  
Db 32358 CTCCGCTGAGCTCTCCGTTATGCTGAGACTACAGGCGCATACCAATGCTGGCTAA 32302  
Qy 7827 TTTTGTATTTTGTGAGGTACAAATGATCTATTTATTTATTTTATTTTATTTT 7886  
Db 32301 -----TAAATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 32285  
Qy 7887 GTAGAGTGAAGTTTCAACATGTTGGCCAGGCTGTCTGCACTCTGACCTGAGTGTAT 7946  
Db 32284 GTAGAGCGGAGTTTCAACATGTTGGCCAGGATGTCTGATCTTGAACCTCA--TGAT 32227  
Qy 7947 CCGCTGCTGCGCTCTCCCAAGTGTGAGATTACAGGATGAGCAACACGCTGGCCC 8006  
Db 32226 CCGCTGCTGCGCTCTCCCAAGTGTGAGATTACAGGATGAGCAACACGCTGGCCC 32167  
Qy 8007 CTGATGATTTTATCTATCTC-----TTGCTCTGCAACATTTGCTGACAGTGG 8061  
Db 32166 CATCCATATTTTGAACCAACCAAGGCAAGCTCTTCTTAATTAATGTTCCCAAGGT 32107  
Qy 8062 ACCATCATGTTCTCCCACTTTCTCACTTCTGATCTTCTGATCTTCTGATCTCAACTG 8121  
Db 32106 GAACAAAGCAAAATTCAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 32047  
Qy 8122 ATATCTCTCTGATGCTCTTTTCTTCTGTAAGATTTTCAAGAGAGAAATCTGAATGCC 8181  
Db 32046 AAAAAAACTACTCTCTTAAATTAATTTAGGGGTTAGTCCATTTATTTATTTTGAAT 31987  
Qy 8182 CCAATCATTTTCTGACCAACCAACATTAAGTGTGATGAGGCTATGATTTGG 8241  
Db 31986 CATTAACCTTAAGCAAAATCTCAGATGTTGAATGTCATTTCCCAAGAGATTGTTATTA 31927  
Qy 8242 CTACATTAATGAGGTTGGAATCTCATCTTACTTCTGACCAAGCAGAT--AGCTC 8298  
Db 31926 GTGTATCATTTAATTTGTCTGCAAAAGTCTCATTTGTTGTTTCTTAATGGCTGTA 31867  
Qy 8299 TGGTTCTCAAAATAGGCGCCCTGGCCAGGTGTGTGCTCATGCTTAATATCCCAAC 8358  
Db 31866 ATCTTTTAATTAACAGATAGAGGCGCAGGCAAGGTGTTCACTCTGTAATCCAGCAC 31807  
Qy 8359 TGTGGAGGCGGAGGGGGGAGATCACTTGAATCCAGAGTTCTAGACCAAGCTGGGCA 8418  
Db 31806 TGTGGAGGCTGAGGAGTGTGATCACTTGAAGCCAGGTGTTCAACCAAGCTGGGCA 31747  
Qy 8419 CATGTGAATCTCATCTCTAATAATAATAATAATAATAATAATAATAATAATAATAATA 8478  
Db 31746 CATGTGAATCTCTCTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 31688  
Qy 8479 CCAATGATCCAGCTGTTCAAGAGGCTGAGGTGAGAGATTTGTCAGATGTGGAGGAG 8538  
Db 31687 TCTGTATATCCCACTTCTTGGAGGCGCAAGGCAATTAAGATTTGTTTAAACCCGCAATCG 31628  
Qy 8539 AGATTCAGTGAACCGTGAATGAGCTCTGCAATCCAGCTGGGTGAGAGATTTGAACCC 8598  
Db 31627 AGGTTCAGAGCAAGAGATTTGCAATTTGCACTCAAGCTGGGTGAGAGATTTGCTC 31568  
Qy 8599 TGTCT 8602  
Db 31567 TGTCT 31564

## RESULT 15

US-10-364-505-9/c

```

; Sequence 9, Application US/10364505
; Publication No. US20030219787A1
; GENERAL INFORMATION:
; APPLICANT: Kere, Juha
; APPLICANT: Taipale, Mikko
; APPLICANT: No. US20030219787A1a-Hemmi, Jaana
; APPLICANT: Kaminen, Nina
; TITLE OF INVENTION: NOVEL HUMAN GENE FUNCTIONALLY RELATED TO DYSLLEXIA
; FILE REFERENCE: 0933-0199P
; CURRENT APPLICATION NUMBER: US/10/364,505
; CURRENT FILING DATE: 2003-02-12
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human DYX1 chromosomal gene region, nucleotides
; OTHER INFORMATION: 150001-20000
US-10-364-505-9

```

```

Query Match      3.4%; Score 344.2; DB 13; Length 50000;
Best Local Similarity 59.3%; Pred. No. 4,1e-77;
Matches 868; Conservative 0; Mismatches 453; Indels 142; Gaps 11;

QY 7186 AAAAAATCCAGGCGCGCGGTGAGTCAAGCTGTAATCTTGAAGCACTTTGGAGGCTG 7245
DB 19952 AAGAGCTAAAGGCGCGGCAAGTGGCTCAACCTGTAAATCCAGAGTTGGAGGCGCA 19893
QY 7246 AGCAGGTGATGCTTGAAGCGGCTGAGCTAGAGTTTGAAGCAGCTTGGGAGCAT 7305
DB 18892 ATGCAAGTGAATCACTGAGGTC-----AGAGTTTGAAGCAGCTTGGCTAAC 19842
QY 7306 GGCAGAACTCTCATCTTCAAAAAATA-TAAAAATTAGTGGGTGTGTGTGTCTTCT 7364
DB 19841 GGTGAAGCCCGTTTCTAATAAAATACAAAAATTGACAGGTGTGTGTGTGTGTCT 19782
QY 7365 ATAGTCCCATCTTCAAGGCTGAG-----CCGGAAGGTGAGG 7406
DB 19781 GTATATCCAGCTACTCCAGGCTGAGGAGAGAAATGCTTGAATCCAGAGGCGCAAG 19722
QY 7407 CTTCAAGGAGCGGAGTCTGCTACTGCTCACTCCAGCTGGGTGAC-AGAGTGAAGCATG 7465
DB 19721 TTGCAAGTGAAGCGAGATCACTGCTCACTCAAGCTTGGGCAAGAGTGAATCTCTG 19662
QY 7466 TCTCAAAAAAACAACAAAA-----AACAAAAACAACAAACAACAAAA 7518
DB 19661 TCTTAAAAAAGAAAAAAGAGAGCATTAATTAAGAAAAAGTTAAAAATATATA 19602
QY 7519 ACCCATATATATATATATATATATATATATATATATATATATATATATATATAT 7578
DB 19601 ACATATATATATATATATATATATATATATATATATATATATATATATATAT 19542
QY 7579 CAACATGACCCAGCTACAGAGTGGGAGGAGCCCTCCCTCTCACTGTGTAAATTTCTTT 7638
DB 19541 AATTATATATATATATATATATATATATATATATATATATATATATATATAT 19482
QY 7639 CTCTGACTCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7698
DB 19481 CATTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19424
QY 7699 ACCCAGGCTGAGTGCATGGGCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7758
DB 19423 GCCCAGGCTGAGTGCATGGGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19364
QY 7759 AAGGATCTCTCTGCTCAGCTCCGCTATATAGTGGAGTACAGGCGCATACCAACATGC 7818
DB 19363 ACGGATCTCTCTGCTCAGCTCCGCTATATAGTGGAGTACAGGCGCATACCAACATGC 19304
QY 7819 CTGGCTAATTTTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7878
DB 19303 CCGGCTAAT-----TTTTTG 19289

```

```

QY 7879 TATTTTAGAGATAGAGGTTTCAACATGTTGGCAGGCTGTCTGAACTCTGACCT 7938
DB 19288 TATTTTAGAGATAGAGGTTTCAACATGTTGGCAGGCTGTCTGAACTCTGACCT 19229
QY 7939 CAGGTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7998
DB 19228 C-GTAAAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19171
QY 7999 CTTGGCCCTCATAGTTTATCTATCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8058
DB 19170 CCGGTGATCATAGTATTTTATATATATATATATATATATATATATATATATATAT 19111
QY 8059 TGGACATCATGTTCTCTCACTTCTCACTCTTCACTCTTCACTCTTCACTCTTCACTCT 8118
DB 19110 TATTTTAGAGATATATATATATATATATATATATATATATATATATATATATAT 19051
QY 8119 CTGATACCTCCCTCAGTGTCTCTTTTCTTCTAGTAAATTTCCAGAGGAGTGTGAAT 8178
DB 19050 CCACAGTACTCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18991
QY 8179 GGCAGTCCATATTTTCAAGACCAACCATTAATAGTGTGATTTCCAGCTTATGTAT 8238
DB 18990 ACTTTGAGGCGGTGAGGAGGAGACCAAGAGTCAAGAGTTCAGAGTTCAGAGCTGAC 18932
QY 8239 TGGCTACATTAATGGGTGGAACTCATCATTTACTTCACTTCACTTCACTTCACTTCA 8298
DB 18931 -----ACATATGTAATCCCT 18915
QY 8299 TGGTTCTCAAAATAGGCGCCCTGCGGCGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8358
DB 18914 CGCTATTAAT-----GCGCGGCGCGGCGGCTCAAGCTGTATCTGACAC 18866
QY 8359 TGTGGAGGCCAGAGGCGGCAATCACTTGAAGTCCAGAGTTTGAAGCAGCTTGGGCA 8418
DB 18865 TTTGGAGGCCAGAGGCGGCAAGCAC--GAGGTCAAGAGATCAAGACCATCTGTGCTAA 18808
QY 8419 CATGTGAATCTCATCTCTCACTTAAATATCAAAAAATTAGCAGGTGTGTGTGTGTGTGT 8478
DB 18807 CACAGTGAACCCGCTCTTCACTTAAATATCAAAAAATTAGCAGGTGTGTGTGTGTGT 18748
QY 8479 CCAGTATCCAGCTGTTCAAGAGGCTGAGTGGAGATTTGCTCGAGTGTGGAGGAG 8538
DB 18747 CCGTATGCTCCAGCTACTTGGAGGCTGAGGAGATTTGCTGTAAATCCGGAAGGCGG 18688
QY 8539 AGATTGCACTGAACCGTGAAGTGTCTGTGAATCCAGGCTGTGTGAAGATTTGAGACCC 8598
DB 18687 AGCTTGACAGGAGCGAGATCCGCGCACTGCACTCAGGCTGGGCAACAGAGCAAGAC 18628
QY 8599 TGTCTCAAAAAACAATAATA 8621
DB 18627 CATCTCAAAAAACAATAATA 18605

```

Search completed: December 16, 2003, 06:45:54  
Job time : 2086.18 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 10:53:38 ; Search time 12512.6 Seconds  
(without alignments)  
19424.068 Million cell updates/sec

Title: US-09-900-448-3\_COPY\_1\_10000

Perfect score: 10000  
Sequence: 1 tccctcccccagcagcgc.....agctgatttcgagcctaac 10000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	664.4	6.6	1201	9	AL534855 AL534855
2	550.4	5.5	552	2	BSM097187 BSM097187
3	493.2	4.9	566	10	BG002737 CVA-CN025
4	436.4	4.4	497	28	AQ138990 HS_3088_B

Result No.	Score	Query Match	Length	ID	Description
5	424	4.2	424	2	BSM089286
6	410.8	4.1	430	2	BSM089310
7	331.4	3.3	537	2	BSM069237
8	287.6	2.9	333	10	BSM27475
9	286	2.9	320	12	BG954333
10	272.2	2.7	6146	28	AQ839831
11	270	2.7	1201	9	AL534854
12	268.8	2.7	1850	11	BC025771
13	265.2	2.7	2296	11	BC035158
14	255	2.5	1850	11	BC025771
15	254	2.5	365	9	AV655253
16	252.4	2.5	2296	11	BC035158
17	251	2.5	985	11	BC035885
18	250	2.5	1201	9	AL534301
19	249.8	2.5	985	11	BC035885
20	248.2	2.5	3828	11	BC035423
21	245	2.5	1201	9	AL534301
22	239	2.4	2551	11	BC024155
23	236	2.4	1034	12	BM806909
24	236	2.4	3777	28	AF101957
25	235.8	2.4	1201	9	AL565216
26	235.6	2.4	420	29	BZ59910
27	235.2	2.4	469	13	BU660220
28	235	2.4	566	9	AL597545
29	234.6	2.3	651	29	AG156348
30	234	2.3	1103	12	BM904488
31	233.6	2.3	378	14	CD516054
32	233.4	2.3	340	28	AQ230483
33	233.4	2.3	505	28	AZ254926
34	233.4	2.3	609	29	AG149236
35	233.4	2.3	873	14	CA454910
36	232.8	2.3	1103	12	BM904488
37	232.6	2.3	690	29	AG105245
38	232.4	2.3	770	13	BX090291
39	231.6	2.3	657	13	BU626531
40	231.4	2.3	437	2	BSM082022
41	231.2	2.3	648	10	BE158921
42	231.2	2.3	1080	13	BQ213854
43	231	2.3	2732	11	BC004387
44	231	2.3	628	28	B54637
45	230.6	2.3	762	28	AQ311398
			493	9	AT733523

## ALIGNMENTS

RESULT 1  
LOCUS AL534855  
DEFINITION AL534855 Homo sapiens PITAL BRAIN Homo sapiens CDNA clone  
ACCESSION AL534855  
VERSION AL534855  
KEYWORDS AL534855.2 GI:30541059  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polyes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12798348.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DP006B110P1&cluster=5958.r Contact :  
Feng Liang Email: fliang@lifetech.com URL:

AL534855 1201 bp mRNA linear EST 12-MAY-2003  
CS0DP006Y122 5-PRIME, mRNA sequence.  
AL534855  
AL534855.2 GI:30541059  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1201)  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12798348.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DP006B110P1&cluster=5958.r Contact :  
Feng Liang Email: fliang@lifetech.com URL:





	Db	241	AATGTGCGTGAATTACTACAAACCAAGTCACACAGAGGGCTCATGATCTTGAGCCCTTGTTA	300
OY		3666	TCTTCTCAGGTTTATCTCTCTCCCTGCACATTACTGTGTGCCAGCATACGAATCTAC	37235
Db		301	TCCTTCAGGTTTATCTCTCCCTGCACATTACTGTGTGCCAGCATACGAATCTAC	360
OY		3726	ATGAGTTTGAGACAACCTGCTTCCCTCATGTTTGGGCTGTGCATGCTCCCTCGCTGG	37859
Db		361	ATGAGTTTGAGACAACCTGCTTCCCTCATGTTTGGGCTGTGCATGCTCCCTCGCTGG	420
OY		3786	TAAACCCCTTTCCTCACTTGTCAAACCTTGAAAATTCCTGTGATTTTCACTCTTGGGC	38454
Db		421	TAAACCCCTTTCCTCACTTGTCAAACCTTGAAAATTCCTGTGATTTTCACTCTTGGGC	480
OY		3846	CCAATGCTTCTCTTGGTGTGAAACCCTTCCACAACTTCTTAGGACACTTAGGCACTC	39050
Db		481	CCAATGCTTCTCTTGGTGTGAAACCCTTCCACAACTTCTTAGGACACTTAGGCACTC	540
OY		3906	TGCTATATTTCT	3917
Db		541	TGCTATATTTCT	552
RESULT 3				
LOCUS		BG002737	566 bp	mRNA linear EST 24-JAN-2001
DEFINITION		QV4-GN0250-171100-549-d12 GN0250 Homo sapiens cDNA, mRNA sequence.		
ACCESSION		BG002737		
VERSION		BG002737.1	GI:12442358	
KEYWORDS		EST.		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Makarewicz, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
RAUTHORS		1. (bases 1 to 566) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
JOURNAL MEDLINE PUBMED COMMENT		2020263 10737800		
		Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?rl=QV4&t=QV4-GN0250-171100-549-d12&t3=2000-11-17&t4=1) Seq primer: puc 18 forward High quality sequence stop: 10 High quality sequence stop: 566. Location/Qualifiers 1..566 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_id="GN0250" /note="Organ: placenta normal; Vector: puc18; Site_1: Small ; Site_2: SmaI; A mini-library was made by cloning products derived from ONESIES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse		
FEATURES				
Source				

BASE COUNT	127	a	147	c	123	g	169	t
Query Match	4.9%	Score 493.2;	DB 10;	Length 566;				
Best Local Similarity	97.8%	Pred. No. 0.37;						
Matches 532;	Conservative 0;	Mismatches 8;	Indels 4;	Gaps 3;				
transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."								
ORIGIN								
Query	6209	CCTCTGAGCTCTCCCTGCTGCTCCCTGAGACCTGTATGTCTCTGACCTCTTGCCTCTT	6268					
Db	26	CCCTCTGAGCTCTCCCTGAGCTCC--GGAACCTCTATGTGTCTGTGT--CTCTCTTGCCTCTT	82					
Qy	6269	TGTTCTTGCAATATCTTTCTGAGTCTCTTGCTGCCCTGATTTATCTCTAGAACTCCATC	6328					
Db	83	TGTTCTTGCAATATCTTTCTGAGTCTCTTGCTGCCCTGATTTATCTCTAGAACTCCATC	142					
Qy	6329	TTGTTTCAAGGTTCCCTGTTGCTTATGTCACAGCCCTGGGAGATAGACCTGCTGGGAGATGA	6388					
Db	143	TTGTTTCAAGGTTCCCTGTTGCTTATGTCACAGCCCTGGGAGATAGACCTGCTGGGAGATGA	202					
Qy	6389	GATGTTCTCAATGCTGAGAACCAAGCTGAGAAAGTGGGTACTTTAGACCTTTAGAGGCT	6448					
Db	203	GATGTTCTCAATGCTGAGAACCAAGCTGAGAAAGTGGGTACTTTAGACCTTTAGAGGCT	262					
Qy	6449	GGCTTCACTAGCTCTGGAAGTCTCTCTGTAGTAGCCAAATGAGATACCCCTCCCTTG	6508					
Db	263	GGCTTCACTAGCTCTGGAAGTCTCTCTGTAGTAGCCAAATGAGATACCCCTCCCTTG	322					
Qy	6509	ACCGTGGGCATCAATGGGTAAAGCCATCTAATATATCTAGGGCTGTCTGAGTTGAGT	6568					
Db	323	ACCGTGGGCATCAATGGGTAAAGCCATCTAATATATCTAGGGCTGTCTGAGTTGAGT	382					
Qy	6569	CAGGCACTAATATAGTCACTGCTGACAGTTGAGAAATATCCCAAGAGAGTGAACAAC	6628					
Db	383	CAGGCACTAATATAGTCACTGCTGACAGTTGAGAAATATCCCAAGAGAGTGAACAAC	442					
Qy	6629	ATCATATCCAACTGAGATATATGATATATTAGACAGTGGTAAAGAAATATATAATCTGA	6688					
Db	443	ATCATATCCAACTGAGATATATGATATATTAGACAGTGGTAAAGAAATATATAATCTGA	502					
Qy	6689	AAATAT--TTTTTTCACACAAATTTTTTGGCTCCCTGACCCCTGGACAAATTTGACAGT	6747					
Db	503	AAATATCTCTTTTCACACAAATTTTTTGGCTCCCTGACCCCTGGACAAATCTGACAGT	562					
Qy	6748	TATG 6751						
Db	563	TATG 566						
RESULT 4								
LOCUS	A0138990	497 bp	DNA	linear	GSS 24-SEP-1998			
DEFINITION	HS 3088_B2_C01_MR CIT Approved Human Genomic Sperm Library D Homo							
LOCUS	A0138990							
ACCESSION	A0138990							
VERSION	A0138990.1	GI:3529643						
KEYWORDS	GSS.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens.							
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
	1 (bases 1 to 497)							
	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,							
	Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and							
	Hood,L.							
	Sequence-tagged connectors: A sequence approach to mapping and							
	scanning the human genome							
	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)							
	99380589							
	10449764							
	Contact: Mahairas GG, Wallace JC, Hood L							
	COMMENT							

High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3088 row: F column: 2  
Class: BAC ends  
High quality sequence stop: 497.

## FEATURES

Location/Qualifiers

1. 497  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=3088 Col=2 Row=F"  
/sex="male"  
/clone\_1ib="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pBelOBAC11, BAC Clones in E-Coli DH10B"

BASE COUNT 92 a 173 c 81 g 144 t 7 others

## ORIGIN

Query Match 4.4%; Score 436.4; DB 28; Length 497;  
Best Local Similarity 93.3%; Pred. No. 3.8;  
Matches 463; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

5550 ATCCACTTTCCCTGAGCTTGTGATCTCAAGTCTCCAGTCTCACTTAACTCCGTT 5609

1 ATCCACTTTCCCTGAGCTTGTGATCTCAAGTCTCCAGTCTCACTTAACTCCGTT 60

5610 GCGACACCTTGAGCCCTTAATCTAGCCCATTTTCATTTGATTTTCCATGCGCCCAT 5669

61 GCGACACCTTGAGCCCTTAATCTAGCCCATTTTCATTTGATTTTCCATGCGCCCAT 120

5670 ATGGGAAACCAACACCCCACTAACCCAGCACTCTTTCACCTTGAACCTCACTTGA 5729

121 ATGGAAACCAACACCCCACTAACCCAGCACTCTTTCACCTTGAACCTCACTTGA 180

5730 CCTCTGCGCTCTTCTGTCTGTTCTCTCAACCATTTCTCTCCAGCCATGAGACAGGA 5789

181 CCTCTGCGCTCTTCTGTCTGTTCTCTCAACCATTTCTCTCCAGCCATGAGACAGGA 240

5790 ATGGGACCTTGAGCCCACTAACCCAGCACTCTTTCACCTTGAACCTCACTTGA 5849

241 ATGGGACCTTGAGCCCACTAACCCAGCACTCTTTCACCTTGAACCTCACTTGA 300

5850 ATCTAGCTTGTCTGCACTGACGTCTGACAAACAGATGTCACCTATGCTTCACTG 5909

301 ATCTAGCTTGTCTGCACTGACGTCTGACAAACAGATGTCACCTATGCTTCACTG 360

5910 AGAGATGCCCCCACTGCCCAATGCTCTACATCTCTTAACTTGAATCCATCC 5969

361 AGAGATGCCCCCACTGCCCAATGCTCTACATCTCTTAACTTGAATCCATCC 420

5970 TTGACACATTTTCATCTGATCTGATCACTGATCACTTAACTTGAATCCATCC 6028

421 TTGACACATTTTCATCTGATCTGATCACTGATCACTTAACTTGAATCCATCC 480

6029 CTCTGATGCGCTTC 6044

481 CTCTGATGCGCTTC 496

RESULT 5  
HSM089286/c standard; RNA; EST; 424 BP.

AC BX496061;

SV BX496061.1

XX 09-MAY-2003 (Rel. 75, Created)

09-MAY-2003 (Rel. 75, Last updated, Version 1)

Homo sapiens mRNA; EST DKFZp799A1126\_r1 (from clone DKFZp799A1126)

EST; expressed sequence tag.

Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Butleria; Primates; Catarrhini; Homnidae; Homo.

[1]

RP 1-424

RA Bioecker H., Boecker M., Mewes H.W., Weil B., Amlid C., Osanger A., Fobo G.,

RA Han M., Wiemann S.;

RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.

RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

CC This is the 5' sequence of the clone insert

CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

CC Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

CC sequenced by GSF (National Research Centre for Biotechnology

CC Ltd., Braunschweig/Germany) within the cDNA sequencing

CC consortium of the German Genome Project.

CC No 51 sequence available.

CC This clone (DKFZp799A1126) is available at the RZPD in Berlin.

CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

Db      64 CAGCCTCTTTCACCTTGACCTCAGCTCTGAGCTCTGCTCTGCTCTGCTCTC 5
QY      5757 ACCC 5760
Db      4 ACCC 1

RESULT 6
HSM089310/C
ID      HSM089310 standard; RNA; EST; 430 BP.
XX      BX496085;
XX      BX496085.1
SV      BX496085.1
DT      09-MAY-2003 (Rel. 75, Created)
DT      09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX      Homo sapiens mRNA; EST DKFZp79C1326_r1 (from clone DKFZp79C1326)
XX      EST; expressed sequence tag.
XX      Homo sapiens (human)
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Primates; Catarrhini; Homiidae; Homo.
XX      [1]
RP      1-430
RA      Bloecher H., Boecher M., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RT      Han M., Wiemann S.;
RL      Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
XX      MIPs, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX      This is the 5' sequence of the clone insert
CC      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC      Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC      sequenced by GFP (National Research Centre for Biotechnology
CC      Ltd., Braunschweig/Germany) within the cDNA sequencing
CC      Consortium of the German Genome Project.
CC      No s1 sequence available.
CC      This clone (DKFZp79C1326) is available at the RZPD in Berlin.
CC      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC      14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1. .430.
FT      /db_xref="taxon:9606"
FT      /mol_type="mRNA"
FT      /organism="Homo sapiens"
FT      /clone="DKFZp79C1326"
FT      /clone_1b="779 (synonym: hnccl). Vector pSport1_sfi; host
FT      DH10B; sites SfiI + SfiIb"
FT      /dev_stage="fetal"
FT      /tissue_type="liver"
XX      Sequence 430 BP; 111 A; 97 C; 142 G; 80 T; 0 other;
SQ

Query Match      4.1%; Score 410.8; DB 2; Length 430;
Best Local Similarity 97.2%; Pred. No. 12;
Matches 418; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

Db      310 AGAGTGCTCAAGGTAACCCGCGAGATGTCGAGACTTTCATGCTGCTGCGCAG 251
QY      5511 AGGTAGAAAGCCCTAGACCTTGAGACCTGTCAGAAATTCATCATTCCCTGAGCTGT 5570
Db      250 AGGTAGAAAGCCCTAGACCTTGAGACCTGTCAGAAATTCATCATTCCCTGAGCTGT 191
QY      5571 GGATCTCAGGTGCTAGCTCTCATCTTAACTCCGTGTGCGACACCTTGAGCTTATC 5630
Db      190 GGATCTCAGGTGCTAGCTCTCATCTTAACTCCGTGTGCGACACCTTGAGCTTATC 131
QY      5631 TAGCCCATTTTCATTGGAATTTTCCATTTGCGCTCATATGAGGAACCCACACCCAC 5690
Db      130 TAGCCCATTTTCATTGGAATTTTCCATTTGCGCTCATATGAGGAACCCACACCCAC 71
QY      5691 TAACCCAGCCATCTTTCACCTTGAGACCTGACCTGACCTGCGCTCTGCTGCTT 5750
Db      70 TAACCCAGCCATCTTTCACCTTGAGACCTGACCTGACCTGCGCTCTGCTGCTT 11
QY      5751 CTCCTCACC 5760
Db      10 CTCCTCACC 1

RESULT 7
HSM069237
ID      HSM069237 standard; RNA; EST; 537 BP.
XX      BX479451;
XX      BX479451.1
SV      BX479451.1
DT      09-MAY-2003 (Rel. 75, Created)
DT      09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX      Homo sapiens mRNA; EST DKFZp686C16211_r1 (from clone DKFZp686C16211)
XX      EST; expressed sequence tag.
XX      Homo sapiens (human)
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Primates; Catarrhini; Homiidae; Homo.
XX      [1]
RP      1-537
RA      Koehner K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RT      Han M., Wiemann S.;
RL      Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
XX      MIPs, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX      This is the 5' sequence of the clone insert
CC      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC      Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC      sequenced by BMRZ (Biomedical Research Center at the
CC      Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA
CC      sequencing consortium of the German Genome Project.
CC      No s1 sequence available.
CC      This clone (DKFZp686C16211) is available at the RZPD in Berlin.
CC      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC      14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1. .537
FT      /db_xref="taxon:9606"
FT      /mol_type="mRNA"
FT      /organism="Homo sapiens"
FT      /clone="DKFZp686C16211"
FT      /clone_1b="686 (synonym: hlccl). Vector pSport1_sfi; host
FT      DH10B; sites SfiI + SfiIb"
FT      /dev_stage="adult"
FT      /tissue_type="cDNA-collection"
XX

```

SQ Sequence 537 BP, 142 A; 110 C; 169 G; 116 T; 0 other;

Query Match 3.3%; Score 331.4; DB 2; Length 537;  
Best Local Similarity 99.7%; Pred. No. 2.2e+02;  
Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCCTCTCCCGAGGAGCCGAGCAAAATCTGATGAGATTCAGACAGGATTCAGACAGCTG 60  
DB TCCCTCTCCCGAGGAGCCGAGCAAAATCTGATGAGATTCAGACAGGATTCAGACAGCTG 264  
QY 61 AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 120  
DB AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 324  
QY 121 TAAGTTTCAGAGGCTGAGAGGAGAAAGATGAGGGGAGCTTGAATAGTGCTC 180  
DB TAAGTTTCAGAGGCTGAGAGGAGAAAGATGAGGGGAGCTTGAATAGTGCTC 384  
QY 325 TAAGTTTCAGAGGCTGAGAGGAGAAAGATGAGGGGAGCTTGAATAGTGCTC 384  
DB 325 TAAGTTTCAGAGGCTGAGAGGAGAAAGATGAGGGGAGCTTGAATAGTGCTC 384  
QY 181 CCATTGCCCCAACCCAGAAAGAGAGATGCGCTGCAATGGGAGAGTGAATAGAGA 240  
DB CCATTGCCCCAACCCAGAAAGAGAGATGCGCTGCAATGGGAGAGTGAATAGAGA 444  
QY 241 CATTGGCTGTAGCAGCGATGCGATTCGCCAGGCTGCCAGAGACTCAGAGAGTCCAGCTT 300  
DB CATTGGCTGTAGCAGCGATGCGATTCGCCAGGCTGCCAGAGACTCAGAGAGTCCAGCTT 504  
QY 301 GCCCACTGACCTATGAGAGGAGATGATGTTC 333  
DB GCCCACTGACCTATGAGAGGAGATGATGTTC 537

RESULT 8  
BF827475 393 bp mRNA linear EST 13-JAN-2001  
LOCUS CM4-HN0022-251100-461-h07 HN0022 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF827475  
VERSION BF827475.1 GI:12171494  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
1 (bases 1 to 393)

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

FEATURES  
source 1. .393  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HN0022"  
/note="Organ: head, normal. Vector: puc18; Site 1: Smal; Site 2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 99 a 81 c 120 g 93 t  
ORIGIN

Query Match 2.9%; Score 287.6; DB 10; Length 393;  
Best Local Similarity 98.4%; Pred. No. 1.6e+03;  
Matches 301; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1 TCCCTCTCCCGAGGAGCCGAGCAAAATCTGATGAGATTCAGACAGGATTCAGACAGCTG 60  
DB TCCCTCTCCCGAGGAGCCGAGCAAAATCTGATGAGATTCAGACAGGATTCAGACAGCTG 121  
QY 61 AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 120  
DB AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 181  
QY 122 AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 181  
DB 122 AAGACAAGTTGTTGAGGAAATTCCTGATGAGATTCATGGGGTCTCAGAGGAGAAATA 181  
QY 121 TAAGTTTCAGAGGCTGAGAGGAGAAAGATG-AGGGGAGCTTGAATAGTGCTC 179  
DB 182 TAAGTTTCAGAGGCTGAGAGGAGAAAGATGTTTGGAGTCTTGAATAGTGCTC 241  
QY 180 CCATTGCCCCAACCCAGAAAGAGAGATGCGCTGCAATGGGAGAGTGAATAGAGA 239  
DB 242 CCATTGCCCCAACCCAGAAAGAGAGATGCGCTGCAATGGGAGAGTGAATAGAGA 301  
QY 240 ACATTGGCTGTAGCAGCGATGCGATTCGCCAGGCTGCCAGAGACTCAGAGAGTCCAGCTT 289  
DB 302 ACATTGGCTGTAGCAGCGATGCGATTCGCCAGGCTGCCAGAGACTCAGAGAGTCCAGCTT 361  
QY 300 TGCCCA 305  
DB 362 TGTCCTCA 367

RESULT 9  
BG954333 320 bp mRNA linear EST 12-JUN-2001  
LOCUS BG954333  
DEFINITION CM4-CT0656-080201-868-h10 CT0656 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG954333  
VERSION BG954333.1 GI:14372504  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
1 (bases 1 to 320)

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br



QY 7897 GATTTCACCAATGTTGGCCAGGCTGTCTCGAATCTCTGACTGATGATCCGCTGCT 7956  
 DB 1602 GATTTCACCTGTGTAGGCAAGATGTCTTCATCTCTGACTC--GTGATCCGCTGCT 1545  
 QY 7957 CGGCTCTCCCAAGTGTCTGATTAAGGATGACCAACCAAGCTGCTGCTC 8005  
 DB 1544 CGGCTCTCCCAAGTGTCTGATTAAGGATGACCAACCAAGCTGCTGCTC 1496  
 RESULT 11  
 AL534854/c 1201 bp mRNA linear EST 12-MAY-2003  
 LOCUS AL534854 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CS0DF006Y122 3-PRIME, mRNA sequence.  
 ACCESSION AL534854  
 VERSION AL534854.2 GI:30541057  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12798347.  
 CONTACT: Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seq@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library: seq@genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 5958.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DF006B11NP1&cluster=5958.r. Contact :  
 Feng Liang Email: fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DF006B11NP1.  
 FEATURES  
 source  
 1.1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF006Y122"  
 /issue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_id="Homo sapiens FETAL BRAIN"  
 /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 BASE COUNT 295 a 266 c 359 g 231 t 50 others  
 ORIGIN  
 Query Match 2.7%; Score 270; DB 9; Length 1201;  
 Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
 Matches 285; Conservative 4; Mismatches 10; Indels 1; Gaps 1;  
 QY 5608 TTGGGACACCTTGGCCCTTAATCTAGCCCAATTCATTCGATTTTCCCATTCGCTC 5667  
 DB 1004 KTKGACACCTTGGCCCTTAATCTAGCCCAATTCATTCGATTTTCCCATTCGCTC 945  
 QY 5668 ATATGGGAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5727  
 DB 944 ATAT-GGAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 886  
 QY 5728 GACCTCTGGGCTCTTCTGTGTTCTCTGACCAATTTCTCTCTCAGGCAAGGACAG 5787  
 DB 885 GACCTCTGGGCTCTTCTGTGTTCTCTGACCAATTTCTCTCTCAGGCAAGGACAG 826  
 QY 5788 GAATGGACTGGCCATGGAAACAGTACCAACATGGCCCTGAGTATATGCGTGTAGCCC 5847

DB 825 GAATGGACTGGCCATGGAAACAGTACCAACATGAGCCCTGAGTATATGCGTGTAGCCC 766  
 QY 5848 ACATCTAGTTTCTGCACTGACGTCTGACCAACCAACCAACCAACCAACCAACCA 5907  
 DB 765 ACATCTAGTTTCTGCACTGACGTCTGACCAACCAACCAACCAACCAACCAACCA 706  
 RESULT 12  
 BC025771/c 1850 bp mRNA linear HTC 04-MAR-2003  
 LOCUS BC025771 Homo sapiens, clone IMAGE:5210159, mRNA.  
 DEFINITION BC025771  
 ACCESSION BC025771.1 GI:22213018  
 VERSION HTC.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1850)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
 CONTACT: MGC help desk  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: http://www.nisc.nih.gov/  
 Contact: nisc.mgc@hgtl.nih.gov.  
 Ahner, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breun, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hight, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Krong, P., Lurie, P., Legeay, R.,  
 Maduro, Q.L., Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W.,  
 Turegion, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, B.D.  
 Clome distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov  
 Series: IRAC Plate: 49 Row: 3 Column: 19  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 20070297  
 This clone has the following problem: retained intron.  
 FEATURES  
 source  
 1.1850  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5210159"  
 /issue\_type="Lung, Spleen, fetal, pooled"  
 /clone\_id="NIH MGC\_122"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 BASE COUNT 524 a 386 c 466 g 474 t  
 ORIGIN  
 Query Match 2.7%; Score 268.8; DB 11; Length 1850;  
 Best Local Similarity 60.5%; Pred. No. 1e+03;  
 Matches 632; Conservative 0; Mismatches 322; Indels 90; Gaps 8;  
 QY 7630 TTTTCTTCTGCACTACAGTTTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 7689  
 DB 1787 TTTTCTTCTGCACTACAGTTTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1728

```

QY 7690 CACTGTGACCCAGGCTGAGTGAATGAGCGCAATCTTGTTCACTGCACTCTGCT 7749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1727 CACTGTGACCCAGGCTGAGTGAATGAGCGCAATCTTGTTCACTGCACTCTGCT 1668
QY 7750 CTTGGGTTCAAGGATCTCTGCTCCTGAGCTCCGCTATAGCTGGACT-ACAGCGCAT 7808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1667 CCGAGGTTCAAGCAATCTCATGCTCCTGAGCTCTGAGTGTGATTAACAGGGAAG 1608
QY 7809 ACCACATGCTGGCTGAATTTTGTATTTTGGGTTACATGATCTATTTAAT 7868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1607 ACTACACAGCTGGC----- 1593
QY 7869 TTAATTTTGTATTTTGTATGATAGATGAGGTTTCAATATGAGCGAGCTGCTCGAA 7928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1592 -TAACTTTTGTATTTTGTATGATAGATGAGGTTTCAATATGAGCGAGCTGCTCGAA 1534
QY 7929 CTCTGACCTGAGGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1533 CTCTGACCTGAGGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1474
QY 7989 AGCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1473 TGCATCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1414
QY 8036 TCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1413 ATTCCTGATATGAGGACGTTAGTATACATAGATGAGGTTGAAACAGAACCTTT 1354
QY 8076 TCCACTTTCCTGACTTATGATCTTTCAGTCTGAGTTCAGATGATGATCTTCTCAGT 8135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1353 GGGTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1294
QY 8136 TGGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1293 TCTTAAACCATCTCTTAAATTTTCACTTCCAGCCCAATGAGGCTCTATCTGAGG 1234
QY 8196 CAGACCA-----CCACATTAAGTGTGATGAGGCTGAGCTGATGATGCTACATTA 8249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1233 CTATCGAATTAATTTTCACTGATATTTTATGATGATGATGATGATGATGATGATG 1174
QY 8250 ATGGGTTGGGAACTCATTTTATTTTCACTTTCAGCAAGAGAGATGCTGCTGCTCA 8309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1173 CTCTGATGATTAATTAATCTCTGCAAGCTTGTCTAA-TTCTGCAATAGCCCTCA 1115
QY 8310 ATAGGCTCCCTGGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1114 AATCCCATATACCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
QY 8370 GAGGGGGGAGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1054 GAGGTGGGAGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995
QY 8430 CTGATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 994 CCGCTCTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 937
QY 8490 AGCTGTTCAAGAGGCTGAGTGTGAGGATGCTGAGTGTGAGGAGAGATGAGTGTG 8549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 936 AGCTTACTGGGAGATGAGGAGAGATGCTGAGGAGAGAGAGAGAGAGAGAGAGAG 877
QY 8550 AACCGGAGCTGAGGCTGAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 876 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
QY 8609 AACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 816 GAGAAACAACTAACAAAAAA 793

```

```

RESULT 13
LOCUS BC035158 2296 bp mRNA linear HTC 04-MAR-2003
DEFINITION Homo sapiens, clone IMAGE:525532, mRNA.

```

```

ACCESSION BC035158
VERSION BC035158.1 GI:23272906
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2296)
AUTHORS Strausberg, R.
TITLES Direct Submission
JOURNAL Submitted (31-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Tohshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAK Plate: 73 Row: 1 Column: 12
This clone has the following problem: retained intron.
FEATURES
source 1..2296
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:525532"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 623 a 575 c 568 g 530 t
ORIGIN
Query Match 2.7%; Score 265.2; DB 11; Length 2296;
Best Local Similarity 60.0%; Pred. No.1e+03;
Matches 608; Conservative 0; Mismatches 308; Indels 98; Gaps 6;
QY 7657 TTGTTGTTGCTGCTGTTGTTGATGAGTGCATCTGTCAACCCAGGCTGAGTCAA 7716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1298 TTTCTTTTCTTTTCTTTTCTTCCAGACAGTCTCACTTTGTCAACCGGCTGAGTACG 1357
QY 7717 TGGCGAATCTTGTTCACTGCACTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 7776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1358 TGGCAATCTGGGCTCACTGCACTCTGCTCTCCAGGTTCAAGATTTCTCTGCTC 1417
QY 7777 AGCTCCGATAGCTGGAGCTACAGGCGCATACCACTGCTGCTGCTGCTGCTGCTG 7836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1418 AGCTCCCGGATAGCTGGAGTTACAGGACAGTGGCACACGCTGCG----- 1464
QY 7837 TTTTGGGTTTACATGATGATTAATTTAATTTTGTATTTAGTAGAGATAG 7896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1465 -----TAATTTTGTATTTAGTAAGATAG 1491
QY 7897 GGTTCACATGTTGGCCAGGCTGCTGCACTCTGCACTCAGGTGATCCGCTGCT 7956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1492 GGTTTACATATTTGGCCAGGCTGATCTTGAATCTGTAAGTGTGATGCTGCTGCT 1551
QY 7957 CGGCTCCCAAGTGTGAGATTAACAGGATGACACAGCGCTGCGCTCTATAGTT 8016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1552 CAGCTCCCAAGCGCTGATTAACAGGAGGTAGAGTACCGTCCCGCTGTGAGCT 1611
QY 8017 TTATATATTCCTT--TCTCTTCAAACTTTGGCTTGAAGTGAACATCATGTTCT 8074

```







```

Db      1349 AACCCAAAGTTCTGTTTCAACCTGATCTATTGTATTAATAACGTCCTCAATPACA 1408
QY      8274 TTGATTTGCANAAAGCAGCATAGCTCTGTGTTCTCAAAATAGAGCCCTCGGCGCAGGTG 8333
Db      1409 GGAATGAGAAAGTGGAGAGAAAGGAGATTTGATAGATGTGTAGAGCAGCAGCCGGCATCA 1468
QY      8334 TGGCTATGCTTAATATCCCAACACTGTGGAGAGCCGAGAGGGGAGATCACTTGAATCC 8393
Db      1469 TGGCACCACCTGTGTATCCAGCATTTGGAGAGCCGAGAGCCGAGATCACTGAGGTC 1528
QY      8394 AGGAGTTCTAGACACAGCTGCGGCAACATGTGAATCTCATCTACTAATAAATACAAA 8453
Db      1529 AAGAGTTGTATACAGCTGCGCAACATAGTGAACCTGTCTCTACTAATAAATACAAA 1588
QY      8454 AATTAGCCAGGTGTGTGTCATGACCAAGTATGCCAGCTGTTGAGAGGCTGAGTGG 8513
Db      1589 GTTAGCCAGGCGTGTGTGTGTGTCCTCTGTTAATACAGCTACTCAGAGGCTGAGGATG 1648
QY      8514 AGGATTGCTGAGTGTGGAGGAGAGATTCAGTGAACCGTGACGTGCTGCAATC 8573
Db      1649 AGAATGCTTGAACCTGGAGGTGTGAGGTGCAGTGAGCCAAATATGCGCACTGCACTC 1708
QY      8574 CAGCCTGGTGAAGATTGAGACCTGTCTCAAAAAAACAATTAATTAATTAAT 8633
Db      1709 CAGCCTGGCTACTGAGTGAAGTCTGTCTCAAAAAAACAATTAATTAATTAAT 1768
QY      8634 A 8634
Db      1769 A 1769

```

## RESULT 15

```

LOCUS      AV655253          365 bp      mRNA      linear      EST 15-JAN-2002
DEFINITION AV655253      GLC Homo sapiens cDNA clone GICBED03 3', mRNA sequence.
ACCESSION  AV655253
VERSION     AV655253.1      GI:9876267
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

```

```

REFERENCE   Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 365)
            Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
            Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
            Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
            Hu, G., Gu, J., Chen, Z. and Han, Z.

```

```

TITLE       Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
            Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

```

## JOURNAL

```

MEDLINE     21625106
PUBMED      11752456

```

## COMMENT

```

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers

```

## FEATURES

## source

```

1..365
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GICBED03"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="adult"
/lab_host="SOLR"
/clone_id="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

```

```

BASE COUNT      69 a      126 c      83 g      87 t
ORIGIN
Query Match      2.5%; Score 254; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 6.3e+03;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      5654 TTCCCATTCCTCCATATGAGGAGAAACCCACACCCCACTAACCAGCCATCTCTCCACC 5713
Db      1 TTCCCATTCCTCCATATGAGGAGAAACCCACACCCCACTAACCAGCCATCTCTCCACC 60
QY      5714 TTGACCTGACTGTGACTCTGGCCCTCTTGTGTGTTCTGCTGACCCATTTCTCTCCA 5773
Db      61 TTGACCTGACTGTGACTCTGGCCCTCTTGTGTGTTCTGCTGACCCATTTCTCTCCA 120
QY      5774 GGCATGACACAGGAATGGAGCTGGCCATGGGAAAGTAAACCATGATGAGTAT 5833
Db      121 GGCATGACACAGGAATGGAGCTGGCCATGGGAAAGTAAACCATGATGAGTAT 180
QY      5834 ATGCGCTGTAGCCCAATCTAGTCTTGTGTGCACTGACGTGTGAACAACATGATGCCACC 5893
Db      181 ATGCGCTGTAGCCCAATCTAGTCTTGTGTGCACTGACGTGTGAACAACATGATGCCACC 240
QY      5894 TATGCTTCAGTGG 5907
Db      241 TATGCTTCAGTGG 254

```

```

Search completed: December 16, 2003, 05:34:16
Job time : 12524.2 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 06:39:26 ; Search time 8787.02 Seconds

(without alignments)  
17398.312 Million cell updates/sec

Title: US-09-900-448-3\_COPY\_10001\_13737

Perfect score: 3737

Sequence: 1 cccaagagagagagtcagtcac.....acatttcgtgtctctca 3737

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2888711 seqs, 2045481336 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

```

Database :
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sbs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sbs:*
28: em_sy:*
29: em_un:*
30: em_vl:*
31: em_hlg_hum:*
32: em_hlg_inv:*
33: em_hlg_other:*
34: em_hlg_mus:*
35: em_hlg_pin:*
36: em_hlg_rtd:*
37: em_hlg_mam:*
38: em_hlg_vrt:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	3730.6	99.8	191556	9 AC068733	AC068733 Homo sapi
C 2	3729	99.8	112295	9 AC084337	AC084337 Homo sapi
C 3	2449	65.5	148198	2 AC024232	AC024232 Homo sapi
4	618	16.5	629	9 HUMHXM07	M36803 Human hemop
5	573.8	15.4	577	9 HUMHXM06	M36802 Human hemop
6	567.4	15.2	249769	2 AC125227	AC125227 Mus muscu
7	557.8	14.9	1467	2 NMHMPX8	X56830 M.musculus
8	520	13.9	238748	2 AC120746	AC120746 Rattus no
C 9	520	13.9	275982	2 AC131626	AC131626 Rattus no
10	444.4	11.9	1621	9 BC005395	BC005395 Homo sapi
11	443.2	11.9	1546	9 HUMHXM06	X03048 Huma hemope
12	431.2	11.5	1485	9 HSHPEPR	X02537 Human mRNA
C 13	365	9.8	400	11 G11396	G11396 SHGC-10293
C 14	265.2	7.1	504	6 AX410520	AX410520 Sequence
15	255.6	6.8	1486	4 OCHEMOEX	X16429 O.cuniculus
16	206.4	5.5	1475	10 BC011246	BC011246 Mus muscu
17	205.6	5.5	1470	10 BC011901	BC011901 Mus muscu
C 18	204	5.5	339	6 AX523973	AX523973 Sequence
C 19	204	5.5	339	6 AX522711	AX522711 Sequence
20	199.6	5.3	1528	4 SSU14751	U14751 Sus scrofa
21	198.6	5.3	1516	10 RATHMPX	M62642 Rat (clone
22	191.6	5.1	1430	10 NMU89889	U89889 Mus muscu
23	160	4.3	160	11 G13507	G13507 human STS S
24	136	3.6	631	6 AX525553	AX525553 Sequence
25	136	3.6	689	6 AX526031	AX526031 Sequence
C 26	116.2	3.1	107527	9 AC005018	AC005018 Homo sapi
C 27	116.2	3.1	186349	2 AC073420	AC073420 Homo sapi
C 28	112.8	3.0	60036	2 AC009610	AC009610 Homo sapi
C 29	112.8	3.0	60036	2 AC009610	AC009610 Homo sapi
C 30	112.8	3.0	132003	9 AC138779	AC138779 Homo sapi
C 31	112.8	3.0	163404	2 AC009614	AC009614 Homo sapi
C 32	112.8	3.0	164084	2 AC026580	AC026580 Homo sapi
C 33	112	3.0	1883	9 AK094459	AK094459 Homo sapi
C 34	112	3.0	93724	2 AC008451	AC008451 Homo sapi
C 35	112	3.0	109884	2 AC010481	AC010481 Homo sapi
36	112	3.0	120041	2 AC009038	AC009038 Homo sapi
C 37	112	3.0	178056	9 AC008536	AC008536 Homo sapi
C 38	112	3.0	181926	2 AC008461	AC008461 Homo sapi
C 39	109.2	2.9	92345	9 AC128687	AC128687 Homo sapi
C 40	109.2	2.9	154936	2 AC021252	AC021252 Homo sapi
41	109.2	2.9	182262	2 AC023963	AC023963 Homo sapi
42	109	2.9	104818	2 AC026690	AC026690 Homo sapi
43	109	2.9	159853	9 AC010451	AC010451 Homo sapi
44	108.6	2.9	195986	2 AC116933	AC116933 Papio anu
C 45	108.2	2.9	56631	9 HS702M17	AL034427 Human DNA

# ALIGNMENTS

```

RESULT 1
AC068733/c 191656 bp DNA linear pri 24-MAY-2002
LOCUS AC068733 Homo sapiens chromosome 11, clone RP11-304C12, complete sequence.
DEFINITION AC068733
ACCESSION AC068733
VERSION AC068733.12 GI:19703148
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 191656)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-304C12
JOURNAL Unpublished

```

REFERENCE  
AUTHORS  
2 (Bases 1 to 191656)  
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barra,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Caselle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hago,B., Heatford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczeky,J.,  
Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeters,R.,  
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
Submitted (08-MAY-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS  
3 (Bases 1 to 191656)  
Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barra,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,  
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhngang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Straus,N., Subramanian,A., Talamas,J., Tessfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS  
4 (Bases 1 to 191656)  
Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barra,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,  
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhngang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Straus,N., Subramanian,A., Talamas,J., Tessfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT  
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2002 this sequence version replaced gi:19683105.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

FEATURES  
source  
-----  
Location/Qualifiers  
1. 191656  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="RP11-304C12"  
/clone\_1lb="RP11-304C12 Human Male BAC"  
complement(189..1091)  
/rpt\_family="L2"  
complement(1150..1092)  
/rpt\_family="GN)n"  
complement(1151..1401)  
/rpt\_family="L2"  
complement(1891..2187)  
/rpt\_family="AluUb"  
3141..3165  
/rpt\_family="CA)n"  
complement(3205..3522)  
/rpt\_family="LMC2"  
complement(4393..4591)  
/rpt\_family="L2"  
4624..4932  
/rpt\_family="AluSp"  
4987..5077  
/rpt\_family="MIR"  
complement(5165..6245)  
/rpt\_family="Tigret3b"  
complement(6246..6555)  
/rpt\_family="AluBo"  
complement(6556..6634)  
6940..7136  
/rpt\_family="L2"  
complement(7607..7849)  
/rpt\_family="L2"  
complement(11242..11318)  
/rpt\_family="MIR"  
complement(11385..11572)  
/rpt\_family="MER69A"  
complement(12325..12520)  
/rpt\_family="MIR"  
12536..12706  
/rpt\_family="MIR"  
12905..13529  
/rpt\_family="LMA3"  
13530..13842  
/rpt\_family="AluSp"  
13843..16076  
/rpt\_family="LMA3"  
14517..14528  
/note="<30 qual SNGI region"  
14536..14542  
/note="<30 qual SNGI region"  
14553..14558

```
unreure /note="<30 qual SNGI region"
14582. .14588 /note="<30 qual SNGI region"
unreure 14899. 14904
repeat_region /note="<30 qual SNGI region"
16077. .16386 /rpc_family="Alusx"
repeat_region 16387. .16988 /rpc_family="L1M33"
repeat_region complement (17002. .17075)
repeat_region /rpc_family="MIR"
repeat_region complement (17907. .18207)
repeat_region /rpc_family="Alusx"
repeat_region complement (18218. .18368)
repeat_region /rpc_family="MIR"
18430. .18711 /rpc_family="Alusx"
repeat_region 18712. .18743 /rpc_family="AT_rich"
repeat_region 19184. .19685 /rpc_family="L2"
repeat_region complement (19738. .19892)
repeat_region /rpc_family="L2"
repeat_region 19894. .20155 /rpc_family="Alusx"
repeat_region 20167. .20189 /rpc_family="L1M33"
repeat_region complement (20649. .20961)
repeat_region /rpc_family="Alusx"
repeat_region 21292. .21411 /rpc_family="L2"
repeat_region 21454. .21594 /rpc_family="L2"
repeat_region 21774. .22026 /rpc_family="L1M33"
repeat_region 22048. .22570 /rpc_family="MIR"
repeat_region 22571. .23617 /rpc_family="MIR41A-int"
repeat_region 23618. .23651 /rpc_family="L1M33"

Query Match 99.84; Score 3730.6; DB 9; Length 191656;
Best Local Similarity 99.94; Pred. No. 0;
Matches 3733; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CTCGAAGAGAGGCTCAGTCAAGCTGACGAGAACAGATTGGAATCATCAGCATATAGAT 60
DB 171352 CTCGAAGAGAGGCTCAGTCAAGCTGACGAGAACAGATTGGAATCATCAGCATATAGAT 171253
QY 61 GGCATTTAAAGCCCTGAGCTAGGTGAGATTACCAAGAGAGTGAAGTGAAGAGAGA 120
DB 171292 GGCATTTAAAGCCCTGAGCTAGGTGAGATTACCAAGAGAGTGAAGTGAAGAGAGA 171233
QY 121 GAAGAGGCCCCAAGTGGGGATTCGAATTTAGTATCAGTTGGAAGAAAGATGATC 180
DB 171232 GAAGAGGCCCCAAGTGGGGATTCGAATTTAGTATCAGTTGGAAGAAAGATGATC 171173
QY 181 AAAAAAGATTAAGAGATTAATCTGGAGAGTCAAGTGTCAAGAAAGCCAAATTCAGAAAAA 240
DB 171172 AAAAAAGATTAAGAGATTAATCTGGAGAGTCAAGTGTCAAGAAAGCCAAATTCAGAAAAA 171113
QY 241 GACATTTAAAGAGAGAGAGTGAAGTCAAGTGTCTCTGAGAGTGAAGTCAAGT 300
DB 171112 GACATTTAAAGAGAGAGAGTGAAGTCAAGTGTCTCTGAGAGTGAAGTCAAGT 171053
QY 301 GAGAGAGAGAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360
DB 171052 GAGAGAGAGAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 170993
QY 361 AACAAATTTCAATGAGTGAAGGAGAAACAGAAATGTAATTGAAGAGATTGAAGAAAAAGAC 420
DB 170992 AACAAATTTCAATGAGTGAAGGAGAAACAGAAATGTAATTGAAGAGATTGAAGAAAAAGAC 170933
```

```
QY 421 AAATGGAGCCTAGATTAATTCCTTAATAGTTGTTGAAAAAGAGAGAGAAAAAGCGG 480
DB 170932 AAATGGAGCCTAGATTAATTCCTTAATAGTTGTTGAAAAAGAGAGAGAAAAAGCGG 170873
QY 481 GTGCTAGCCAGCTACTCCCTCACTCTTCACCACTCACTAGAGAGAGAGAGAGAGAGAG 540
DB 170872 GTGCTAGCCAGCTACTCCCTCACTCTTCACCACTCACTAGAGAGAGAGAGAGAGAG 170813
QY 541 AGCCAGAGTGAAGAACTTCAAGTGAAGAGTGTCTCTTTTAAGTTCTGAGACCTGTAT 600
DB 170812 AGCCAGAGTGAAGAACTTCAAGTGAAGAGTGTCTCTTTTAAGTTCTGAGACCTGTAT 170753
QY 601 TTCAATATATTAACCGCATCTCTGACATGACCACTGAAATTCCTTAGAGAGTGCCTGC 660
DB 170752 TTCAATATATTAACCGCATCTCTGACATGACCACTGAAATTCCTTAGAGAGTGCCTGC 170693
QY 661 CAACCCATGATGTTGGCTTCACTGGAACCTTAGCACTGTTTCACACTTGCCTTTC 720
DB 170692 CAACCCATGATGTTGGCTTCACTGGAACCTTAGCACTGTTTCACACTTGCCTTTC 170633
QY 721 TTTCAGGCACTGCTGATTCAGTTTCAGCCAGGAGCAAGTGCCTCAATGCTGACCA 780
DB 170632 TTTCAGGCACTGCTGATTCAGTTTCAGCCAGGAGCAAGTGCCTCAATGCTGACCA 170573
QY 781 GTCTGCTATTTCTCTCTTCACTGAGGCTCTTCCATCTTGGCTCTGAGATGATCT 840
DB 170572 GTCTGCTATTTCTCTCTTCACTGAGGCTCTTCCATCTTGGCTCTGAGATGATCT 170513
QY 841 CTCCTCTCATGACTCATTTCTGATTCATCACTAGCCCTTCTGCTGAGGCTTCTGC 900
DB 170512 CTCCTCTCATGACTCATTTCTGATTCATCACTAGCCCTTCTGCTGAGGCTTCTGC 170453
QY 901 CAGGCGCCTTAAGCACTTATGATTCACAGGAGCCCACTAGGCTCTGACACCC 960
DB 170452 CAGGCGCCTTAAGCACTTATGATTCACAGGAGCCCACTAGGCTCTGACACCC 170393
QY 961 AGCCGGAGTGGCTGATAGCTGAGCCCATCTGATCACTAGGCTCTGAGGCTCTTACAG 1020
DB 170392 AGCCGGAGTGGCTGATAGCTGAGCCCATCTGATCACTAGGCTCTGAGGCTCTTACAG 170333
QY 1021 GTGATGCTGCTCTTCTCTGGAAGAAACTTATCTGCTGAGTGTATGAGGAGG 1080
DB 170332 GTGATGCTGCTCTTCTCTGGAAGAAACTTATCTGCTGAGTGTATGAGGAGG 170273
QY 1081 GAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1140
DB 170272 GAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 170213
QY 1141 TCTTCAGAGGCACTGATATATCTTCTGACAAAGGAGGCTATACCTTATGTAAGG 1200
DB 170212 TCTTCAGAGGCACTGATATATCTTCTGACAAAGGAGGCTATACCTTATGTAAGG 170153
QY 1201 GTTATCCGAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 170152 GTTATCCGAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170093
QY 1261 TGGATGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 170092 TGGATGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170033
QY 1321 TTCTGGGCTCTTGAAGGAGAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 170032 TTCTGGGCTCTTGAAGGAGAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 169973
QY 1381 GGCATGAGAGGCTTATGATCAAGATTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 169972 GGCATGAGAGGCTTATGATCAAGATTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 169913
QY 1441 ATGACATGGAAGCATGATGATGTTGCTCTTCTCCAGAGAGAGAGAGAGAGAGAGAG 1500
DB 169912 ATGACATGGAAGCATGATGATGTTGCTCTTCTCCAGAGAGAGAGAGAGAGAGAGAG 169853
```

QY	1501	GGAGCCTGAAGTCAGAGAGCCCAAGCCACGTGGAACAGAGCTTCTTGCGCCCATGAGAAAGGT	1560
Db	169852	GGACCTGAAGTCAGAGAGCCCAAGCCACGTGGAACAGAGCTTCTTGCGCCCATGAGAAAGGT	169793
QY	1561	AGACGGAGGCTTGATGATGAAAAAGTCCCTTGCGCCCTAACTCAATGTTCCGCCAATGATGCC	1620
Db	169792	AGACGGAGGCTTGATGATGAAAAAGTCCCTTGCGCCCTAACTCAATGTTCCGCCAATGATGCC	169733
QY	1621	CGGCTTGTAACCTGATCCATGATGATCCCAATTTGTACGTGCTACAGATGATGAGAACTGAA	1680
Db	169732	CGGCTTGTAACCTGATCCATGATGATCCCAATTTGTACTGTACAGATGATGAGAACTGAA	169673
QY	1681	TGCAGCAAGGGCCCTTCCGCCAACCCCAAGATGGAACAAGTCTCTGCGCTGCACCTACG	1740
Db	169672	TGCAGCAAGGGCCCTTCCGCCAACCCCAAGATGGAACAAGTCTCTGCGCTGCACCTACG	169613
QY	1741	AGGGGCTTCTGACATGATGATGCTGCGCTGCGCCCACTCTCAATGTTCCATATTAAGACA	1800
Db	169612	AGGGGCTTCTGACATGATGATGCTGCGCTGCGCCCACTCTCAATGTTCCATATTAAGACA	169553
QY	1801	GATTGCTTCTTGCTTCTCACTGAGGGGCTTCTGACATGAGTCTGCGCTGCGCCCACT	1860
Db	169552	GATTGCTTCTTGCTTCTCACTGAGGGGCTTCTGACATGAGTCTGCGCTGCGCCCACT	169493
QY	1861	CCCCAGTTTCTCATATTAAGACAGATTGCTTCTTCACTTGAATCAAGGACCTTGCTG	1920
Db	169492	CCCCAGTTTCTCATATTAAGACAGATTGCTTCTTCACTTGAATCAAGGACCTTGCTG	169433
QY	1921	TGAAACAATCTTCTTCTTGAGTTGAAAAAGTTAGACCTTCTCTTGAGGGGTGAGC	1980
Db	169432	TGAAACAATCTTCTTCTTGAGTTGAAAAAGTTAGACCTTCTCTTGAGGGGTGAGC	169373
QY	1981	TCMAACAAAGCTGTGAGAAACAAAGGAGGGAGCACTAAGGGGCAAACTATCTTGCGC	2040
Db	169372	TCMAACAAAGCTGTGAGAAACAAAGGAGGGAGCACTAAGGGGCAAACTATCTTGCGC	169313
QY	2041	AGATGATTTTGATGGTCCAGATCATTAACCTAAGCTCTTGAGACTATCTACACTAGTGG	2100
Db	169312	AGATGATTTTGATGGTCCAGATCATTAACCTAAGCTCTTGAGACTATCTACACTAGTGG	169253
QY	2101	GGGAAAGAGAACAGAGTGGGAGAGAGAACAGCTGAGTTTATACACAGTAAGAGTGG	2160
Db	169252	GGGAAAGAGAACAGAGTGGGAGAGAGAACAGCTGAGTTTATACACAGTAAGAGTGG	169193
QY	2161	AGCTAGACTCTGATTCAACTTGCTGATGATGGCACAACCAGCGCAGCATCAGA	2220
Db	169192	AGCTAGACTCTGATTCAACTTGCTGATGATGGCACAACCAGCGCAGCATCAGA	169133
QY	2221	AACAACAGGGCTCTGGGGCACTATGATGTGCAGAGAGATTGGCTCAGAGTTTGCGGT	2280
Db	169132	AACAACAGGGCTCTGGGGCACTATGATGTGCAGAGAGATTGGCTCAGAGTTTGCGGT	169073
QY	2281	AGGAGGTCCAATCTGGGGGAGCCCAAAATTATGTTCTGGGTGATTCAAGTAACACACATC	2340
Db	169072	AGGAGGTCCAATCTGGGGGAGCCCAAAATTATGTTCTGGGTGATTCAAGTAACACACATC	169013
QY	2341	ATGCTTGTTGTCATGATGATGATGAGCATGACAAAGTGAATGAAGTTGAATGGGGAAAAA	2400
Db	169012	ATGCTTGTTGTCATGATGATGATGAGCATGACAAAGTGAATGAAGTTGAATGGGGAAAAA	168953
QY	2401	GAATATCACACAGCTGTGTGTACAGAGCAAGCTGAGAGAGAGAGAAAGATGATGGCA	2460
Db	168952	GAATATCACACAGCTGTGTGTACAGAGCAAGCTGAGAGAGAGAGAAAGATGATGGCA	168893
QY	2461	CCATGAGAGCAATTTTGCAAAACACAGTCCCTGGGAATCTTGCTGAGAGCTCAGAGACTTT	2520
Db	168892	CCATGAGAGCAATTTTGCAAAACACAGTCCCTGGGAATCTTGCTGAGAGCTCAGAGACTTT	168833
QY	2521	GCTGGACACAGAGATCTGCGCTACCCAAATTAGCTCTGGGATCTTGCAACCAATCTAGACC	2580
Db	168832	GCTGGACACAGAGATCTGCGCTACCCAAATTAGCTCTGGGATCTTGCAACCAATCTAGACC	168773
QY	2581	AGCAATATGCACTGGCAGAGAGATTGCAAGTGTCTTGTTATTTTCTGTGATTAACCTGGTG	2640

Db	168772	ACCAATGTCACCTGGACAGAGGATTGCACTGCTGGTTATTTCTGCTCATTAACCTGGTG	168713
Qy	2641	AAGCTTTGGGTTCCAAATTTTGCTGACAGCTGTTTAACCTGGAAATTGGGCTTAGACTATA	2700
Db	168712	AAGGCTTTGGGTTCCAAATTTTGCTGACAGCTGTTTAACCTGGAAATTGGGCTTAGACTATA	168653
Qy	2701	GGTACGTATGTCTCAGACAAAGGCCCTATTTCCTCCACTGCTTTTACAACCACTGAGGTT	2760
Db	168652	GGTACGTATGTCTCAGACAAAGGCCCTGTTCCTCCACTGCTTTTACAACCACTGAGGTT	168593
Qy	2761	GGAGGCTGGCTGTTTACGCTCCAAAAAAATAGCTGAGTTTCCAGACAGAGGCCCTTATT	2820
Db	168552	GGAGGCTGGCTGTTTACGCTCCAAAAAAATAGCTGAGTTTCCAGACAGAGGCCCTTATT	168533
Qy	2821	CTGACCTTCCGTGCTTAGCCTCATATTTCTTCTGTAAAAATAGACAAATGCCACCA	2880
Db	168512	CTGACCTTCCGTGCTTAGCCTCATATTTCTTCTGTAAAAATAGACAAATGCCACCA	168473
Qy	2881	CCTTCCAGTGAACAAAGAAATATAGCTCAAAACCCATCCCTTGAACCTGTCTTGGGAAAGGCG	2940
Db	168472	CCTTCCAGTGAACAAAGAAATATAGCTCAAAACCCATCCCTTGAACCTGTCTTGGGAAAGGCG	168413
Qy	2941	TCCTGACGTGAACCCAGACTGGGCTCAATGGGCTCATAGATCTGGAAGCAGGCCCTCC	3000
Db	168412	TCCTGACGTGAACCCAGACTGGGCTCAATGGGCTCATAGATCTGGAAGCAGGCCCTCC	168353
Qy	3001	AACCTGTCAAGCATTGTGCTCCCTAGAGACTTTTATGAGGTGAGATAGACTTAACAACCTCT	3060
Db	168352	AACCTGTCAAGCATTGTGCTCCCTAGAGACTTTTATGAGGTGAGATAGACTTAACAACCTCT	168293
Qy	3061	GACTGTCAACAAGGCTTTGTATCTGGAGGCCAGGCTATATAGTGGCTCAAGCTTAAAG	3120
Db	168292	GACTGTCAACAAGGCTTTGTATCTGGAGGCCAGGCTATATAGTGGCTCAAGCTTAAAG	168233
Qy	3121	GCTGGGAGCTGGGGGACAGTGTCTCAGATTAGGCTCTATGGAAGTTGACTGGAGCTG	3180
Db	168232	GCTGGGAGCTGGGGGACAGTGTCTCAGATTAGGCTCTATGGAAGTTGACTGGAGCTG	168173
Qy	3181	AGAAAGAGGTTAAGGGGCCAAGACAGAGGTTGTGGGTTACTCCTTTAGAGACCTTGA	3240
Db	168172	AGAAAGAGGTTAAGGGGCCAAGACAGAGGTTGTGGGTTACTCCTTTAGAGACCTTGA	168113
Qy	3241	GCTTACTTTTCATTCTTAATGATGTCTTGAATGGCTACCTCAAGGAGTGGCTCTAG	3300
Db	168112	GCTTACTTTTCATTCTTAATGATGTGTGATGGCTACCTCAAGGAGTGGCTCTAG	168053
Qy	3301	TCTAAGGGGTGGAGACAGAGACAGAGTTTCAGGTCTGTGCTCTTATCAAGTTATGACATA	3360
Db	168052	TCTAAGGGGTGGAGACAGAGACAGAGTTTCAGGTCTGTGCTCTTATCAAGTTATGACATA	167993
Qy	3361	CACCTAGGAGCCACTGCTGATCATGACAGAGGAGCTTAAAGGTGTCTAAACAGTTATCCA	3420
Db	167992	CACCTAGGAGCCACTGCTGATCATGACAGAGGAGCTTAAAGGTGTCTTAAACAGTTATCCA	167933
Qy	3421	CAACTGTATATCCCAAGTTAACTTTCTTGTGTTTCAAGGACAGGAGTACTAATGCTC	3480
Db	167932	CAACTGTATATCCCAAGTTAACTTTCTTGTGTTTCAAGGACAGGAGTACTAATGCTC	167873
Qy	3481	CCCTTTCTCTTCTCTCCACGTGTTCTCTTCCAGGAAATCCTCTAGCTTGTCTCAAGG	3540
Db	167872	CCCTTTCTCTTCTCTCCACGTGTTCTCTTCCAGGAAATCCTCTAGCTTGTCTCAAGG	167813
Qy	3541	AACCTCCGAAGATGGTTTGTTCATCTCAGTTAAGGTGCTATPAAGAAATATCTTAGAGT	3600
Db	167812	AACCTCCGAAGATGGTTTGTTCATCTCAGTTAAGGTGCTATPAAGAAATATCTTAGAGT	167753
Qy	3601	GGGTATCTATCAGCAATAGAAATTAATGTTCACAATTTCTGAGGCTGGAAATCCAAAG	3660
Db	167752	GGGTATCTATCAGCAATAGAAATTAATGTTCACAATTTCTGAGGCTGGAAATCCAAAG	167693
Qy	3661	ATCAAGCTCCAGAGGTTCACTGTCTGTGATGCTGTTCTGCTTCAAGATGGCACCT	3720

Db 167692 ATCAAGGCTTCACAGGTTCACTGCTGAGNGCTTCTTCTGCTTCAGAGATGAC 167633

QY 3721 TTTTCTGCTGCTTCTCA 3737

Db 167632 TTTTCTGCTGCTTCTCA 167616

RESULT 2

AC084337/c 112295 bp DNA linear PRI 01-JUN-2002

LOCUS Homo sapiens chromosome 11, clone CTD-2010116, complete sequence.

DEFINITION AC084337

AC084337.7 GI:21306722

VERSION HTG.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 112295)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 11, clone CTD-2010116

JOURNAL Unpublished

2 (bases 1 to 112295)

REFERENCE 2 (bases 1 to 112295)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlino, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Mlenge, V., Morow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olyar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 112295)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Cooke, P., Dearlino, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kama, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mlenge, V., Murphy, T., Naylor, J., Nguyen, C., Nicoll, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunthang, P., Pierre, N., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rieck, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 112295)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Cooke, P., Dearlino, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kama, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mlenge, V., Murphy, T., Naylor, J., Nguyen, C., Nicoll, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunthang, P., Pierre, N., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rieck, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLES

## JOURNAL

Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 1, 2002 this sequence version replaced gi:19683126.

All repeats were identified using RepeatMasker:

Smith, A.P.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RV/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIMR

Web site: <http://www-seg.wi.mit.edu>

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10956

Center clone name: 2010\_I\_16

-----

Only the last 112.3 kb of this clone are being submitted.

The remainder overlaps AC068733 [WIGR project L10266].

Location/Qualifiers

FEATURES

source 1..112295

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="CTD-2010116"

/clone\_11b="CTD1 Human BAC"

repeat\_region complement(949..1264)

/rpt\_family="AluSg"

repeat\_region 1269..1568

/rpt\_family="L2"

repeat\_region complement(2334..2584)

/rpt\_family="AluSg"

repeat\_region complement(2429..2433)

/note="<30 qual SNGL region"

repeat\_region complement(2462..2466)

/note="<30 qual SNGL region"

repeat\_region complement(2478..2503)

/note="<30 qual SNGL region"

repeat\_region complement(2508..2537)

/note="<30 qual SNGL region"

repeat\_region complement(2577..2583)

/note="<30 qual SNGL region"

repeat\_region 2772..3185

/rpt\_family="L2"

repeat\_region complement(4149..4288)

/rpt\_family="L2"

repeat\_region 4268..4443

/rpt\_family="L2"

repeat\_region 4453..4524

```

repeat_region      /rpc_family="L2"      4563. .5531
                    /rpc_family="L2"      complement(5560. .5857)
repeat_region      /rpc_family="AluSx"    5861. .6365
                    /rpc_family="L2"      complement(6378. .7246)
repeat_region      /rpc_family="L2"      7898. .7942
                    /rpc_family="GC_rich"  8243. .8264
repeat_region      /rpc_family="GC_rich"  complement(8355. .8589)
                    /rpc_family="MIR"     complement(8875. .9148)
repeat_region      /rpc_family="AluBo"   9264. .9349
                    /rpc_family="L2"      complement(9361. .9480)
repeat_region      /rpc_family="L1MC5"   complement(9584. .9639)
repeat_region      /rpc_family="L2"      9656. .9676
                    /rpc_family="(TTCA)n" complement(9677. .10005)
repeat_region      /rpc_family="L2"      10006. .10186
                    /rpc_family="MSTA"    10187. .10496
repeat_region      /rpc_family="AluSg"    10497. .10671
                    /rpc_family="MSTA"    complement(10672. .10981)
repeat_region      /rpc_family="AluY"    10982. .11052
                    /rpc_family="MSTA"    complement(11053. .11484)
repeat_region      /rpc_family="L2"      11484. .12337
                    /rpc_family="L2"      complement(12337. .12474)
repeat_region      /rpc_family="FlAM C"  complement(12507. .12954)
                    /rpc_family="L2"      12958. .13714
repeat_region      /rpc_family="L1MA7"   complement(14073. .14372)
repeat_region      /rpc_family="AluSg"   14466. .14974
                    /rpc_family="MLTK"    complement(14975. .15266)
repeat_region      /rpc_family="AluSx"   15267. .15301
                    /rpc_family="MLTK"    complement(15301. .15440)
repeat_region      /rpc_family="LTR50"  15440. .15632
                    /rpc_family="L1B17"  15632. .15817
repeat_region      /rpc_family="LTR33A"  complement(15824. .15945)
repeat_region      /rpc_family="MIR"    16067. .16090
                    /rpc_family="AT_rich" complement(16091. .16371)
repeat_region      /rpc_family="AluBo"

```

Query March	99.84%	Score 3729	DB 9	Length 112295
Best Local Similarity	99.98%	Pred. No. 0		
Matches 3732	Conservative	0	Mismatches 5	Indels 0
			Gaps	0
QY	1	CTCAAGAGAGGCTCATCTAGAGCTGACGAGAAACAGATTGGAAGTCATCAGCATTATAGAT	60	
Db	21765	CTCAAGAGAGAGGCTCATCTAGAGCTGACGAGAAACAGATTGGAAGTCATCAGCATTATAGAT	217068	
OY	61	GGCATTTTAAAGCCCTTGAGCTAGGTGAGATTACCAAGGAAGTGAGAGGTAGAGAGACAGA	120	

Db	21705	GGCATTTAAAGCCCTTGACATCAGTGAATTACCAAGAAAGTGAAGTGAAGAGAA	216448
QY	121	GAAGAGCCCAAGTAGGGGATTCAAATATTTAGATATCAGGTTGAAGAAAAGTAGTC	180
Db	21645	GAAGAGCCCAAGTAGGGGATTCAAATATTTAGATATCAGGTTGAAGAAAAGTAGTC	215868
QY	181	AAAAAATAGAGGAATACGTGGAGAGTCAGGTGCACAGAACCAATTCACAAAAA	240
Db	21585	AAAAAATAGAGGAATACGTGGAGAGTCAGGTGCACAGAACCAATTCACAAAAA	215286
QY	241	GACATTTAAAGAGAAAGTAGTAGAGCAGTCACTGCTCTGAGAGTAGGGTCAGAT	300
Db	21525	GACATTTAAAGAGAAAGTAGTAGAGCAGTCACTGCTCTGAGAGTAGGGTCAGAT	214668
QY	301	GAGAACAGAGATTTGACCATGAGATTTCCGAATTGGAAATATCTAGCAACTGGATAG	360
Db	21465	GAGAACAGAGATTTGACCATGAGATTTCCGAATTGGAAATATCTAGCAACTGGATAG	214068
QY	361	AACAAATTTCAATGTTGAGGGAAACAGAAAGTCATATTTGAAGAGATTTGAGAAAAAGAC	420
Db	21405	AACAAATTTCAATGTTGAGGGAAACAGAAAGTCATATTTGAAGAGATTTGAGAAAAAGAC	213468
QY	421	AAATGGAGAGCTAGATTAATCTCTTAATTAAGTTGTGTGAAGAGAGAGAAAAACGGG	480
Db	21345	AAATGGAGAGCTAGATTAATCTCTTAATTAAGTTGTGTGAAGAGAGAGAAAAACGGG	212868
QY	481	GTGCTAGCCGAGCTACTCCCTCACTCTTCCACCACTCATAGGAGAGACTGAGAACAC	540
Db	21285	GTGCTAGCCGAGCTACTCCCTCACTCTTCCACCACTCATAGGAGAGACTGAGAACAC	212268
QY	541	AGCCAGAGTGAAGACATTCACTAGAAAGTGTGCTCTTTTAATCTTCTGACACTGTAT	600
Db	21225	AGCCAGAGTGAAGACATTCACTAGAAAGTGTGCTCTTTTAATCTTCTGACACTGTAT	211668
QY	601	TTCAATATCTATAACCGCATCTCTGTACATAGGACACTGAATACTTAGGAGAGTCCGC	660
Db	21165	TTCAATATCTATAACCGCATCTCTGTACATAGGACACTGAATACTTAGGAGAGTCCGC	211068
QY	661	CAACCCCATGATGTGGCTTACCTTGAAAATTAGCACTGTGTTTCCACACTTGCCTTTC	720
Db	21105	CAACCCCATGATGTGGCTTACCTTGAAAATTAGCACTGTGTTTCCACACTTGCCTTTC	210468
QY	721	TTTTCAGGACACTGTGTGATTCAGATTTCAAGCAGAGGGACAGTGCACAACTTGCTACCA	780
Db	21045	TTTTCAGGACACTGTGTGATTCAGATTTCAAGCAGAGGGACAGTGCACAACTTGCTACCA	209868
QY	781	GTCATGTCTAATTTCTCCCTCTCAACCGGCTCTTCATCTTGAGCTCTGAGATGCATCT	840
Db	20985	GTCATGTCTAATTTCTCCCTCTCAACCGGCTCTTCATCTTGAGCTCTGAGATGCATCT	209268
QY	841	CTCCCTCTCATGACTCATTTCTGCAATTCACTAGCCCTTCTCTGCTGAGGCTTCTGC	900
Db	20925	CTCCCTCTCATGACTCATTTCTGCAATTCACTAGCCCTTCTCTGCTGAGGCTTCTGC	208668
QY	901	CAGCGGCCCTTAGAGCAACTATGTATTCACAGGGACCACTAGTGGCTCTGACACCC	960
Db	20865	CAGCGGCCCTTAGAGCAACTATGTATTCACAGGGACCACTAGTGGCTCTGACACCC	208068
QY	961	AGCCGGGATGGCTGGGATAGCTGGGCCCATTTGCTCATAGTGGCCCAAGGGTCTTCAGCA	1020
Db	20805	AGCCGGGATGGCTGGGATAGCTGGGCCCATTTGCTCATAGTGGCCCAAGGGTCTTCAGCA	207468
QY	1021	GTGATAGCTGCTTTTCTTGGAGAAAAAACTCTATCTGTCTCAGGTGTGTATTTGGGGGA	1080
Db	20745	GTGATAGCTGCTTTTCTTGGAGAAAAAACTCTATCTGTCTCAGGTGTGTATTTGGGGGA	206868
QY	1081	GAGGCTTAGAGTAGACTGGGCAAGCATATTCAACTCTGTATTTATTTACCATCTTTTG	1140
Db	20685	GAGGCTTAGAGTAGACTGGGCAAGCATATTCAACTCTGTATTTATTTACCATCTTTTG	206268
QY	1141	TCTCTCAGGGCACTCAGGTAATGTCTTCTGTACAAAGGAGGCTATACCTTAGTAAGG	1200



Db 20625 TCCTCCAGGACCCAGATATATGTCTTCTGACAAAGGAGGCTATACCTAGTAAGCG 20565  
QY 1201 GTTATCCGAAGCGCTGAGAGAAAGAGTGGGACCCCTCATGGGATTTATCTGGAATCTG 1260  
Db 20565 GTTATCCGAAGCGCTGAGAGAAAGAGTGGGACCCCTCATGGGATTTATCTGGAATCTG 20506  
QY 1261 TGGATCGGCGCTTATCTGCGCTGGGCTTCTGCGGCTCATATCATGCGAGTGGGCG 1320  
Db 20505 TGGATCGGCGCTTATCTGCGCTGGGCTTCTGCGGCTCATATCATGCGAGTGGGCG 20446  
QY 1321 TTCTGGGTGCTTAAAGAGGACGCTTGTCTGCTACCTGTCTGTGGCATATGCCACAG 1380  
Db 20445 TTCTGGGTGCTTAAAGAGGACGCTTGTCTGCTACCTGTCTGTGGCATATGCCACAG 20386  
QY 1381 GGGATGAAAGGCGCTAGGTCAAGATCCCGAGGCAATGAAAGGCTTAGAGATCCCG 1440  
Db 20385 GGGATGAAAGGCGCTAGGTCAAGATCCCGAGGCAATGAAAGGCTTAGAGATCCCG 20326  
QY 1441 ATGACATGGAAGGCAATGCTATGTTGGTGCCTTCTCCGAGAGCGGCGCTGTGGCT 1500  
Db 20325 ATGACATGGAAGGCAATGCTATGTTGGTGCCTTCTCCGAGAGCGGCGCTGTGGCT 20266  
QY 1501 GGAAGCTGAAGTCAAGAGCCCAAGCCAGTGAAGAGCTTCTTGGCCCATGAAAGT 1560  
Db 20265 GGAAGCTGAAGTCAAGAGCCCAAGCCAGTGAAGAGCTTCTTGGCCCATGAAAGT 20206  
QY 1561 AGACGAGGCTTGTATGAGAAAGTCCCTTGGCCCTTAATCATATGTTCCGCCATGGTCC 1620  
Db 20205 AGACGAGGCTTGTATGAGAAAGTCCCTTGGCCCTTAATCATATGTTCCGCCATGGTCC 20146  
QY 1621 CGGCTTGATCTCATCATGATGCTCCCAATTTGTACTGTGATCAGATGTGAGAAAGTGA 1680  
Db 20145 CGGCTTGATCTCATCATGATGCTCCCAATTTGTACTGTGATCAGATGTGAGAAAGTGA 20086  
QY 1681 TGAAGCCAAAGGCGCTTCCGCAACCCCAAGATGACAGCTCTCTGGGCTGCACTGCTG 1740  
Db 20085 TGAAGCCAAAGGCGCTTCCGCAACCCCAAGATGACAGCTCTCTGGGCTGCACTGCTG 20026  
QY 1741 AGGGGCGCTTGTGACATGAGTGTGGCGCTGCGCCCACTCTTAATGTTCTCATATTAAGACA 1800  
Db 20025 AGGGGCGCTTGTGACATGAGTGTGGCGCTGCGCCCACTCTTAATGTTCTCATATTAAGACA 19966  
QY 1801 GATTGCTTCTGCTCTCACTGAGGGGCGCTTGTGACATGAGTGTGGCGCTGCGCCCACT 1860  
Db 19965 GATTGCTTCTGCTCTCACTGAGGGGCGCTTGTGACATGAGTGTGGCGCTGCGCCCACT 19906  
QY 1861 CCCAGGTTCTCAATTAAGAGAGATGCTTCTTCACTTGAATCAAGGAGCTTGGTGG 1920  
Db 19905 CCCAGGTTCTCAATTAAGAGAGATGCTTCTTCACTTGAATCAAGGAGCTTGGTGG 19846  
QY 1921 TGAACCAATCTTCTTCTTGAAGTTGAAGATGACATTTCTCTTGAAGGCTGTGAGC 1980  
Db 19845 TGAACCAATCTTCTTCTTGAAGTTGAAGATGACATTTCTCTTGAAGGCTGTGAGC 19786  
QY 1981 TCAACCAAGGCTGAGAAACAAAGGAGGAGAGACATTAAGGGGCAACCTATCTGCGC 2040  
Db 19785 TCAACCAAGGCTGAGAAACAAAGGAGGAGAGACATTAAGGGGCAACCTATCTGCGC 19726  
QY 2041 AGATGATCTTGAAGTCAAGATCAATTAAGTACTTCTTGAAGTCAATCACTAGTGG 2100  
Db 19725 AGATGATCTTGAAGTCAAGATCAATTAAGTACTTCTTGAAGTCAATCACTAGTGG 19666  
QY 2101 GGGAAAGAGAACAGAGTGGGAAAGAGAGACAGCTGAGTTTATACAGCAATTAAGAGTGG 2160  
Db 19665 GGGAAAGAGAACAGAGTGGGAAAGAGAGAGACAGCTGAGTTTATACAGCAATTAAGAGTGG 19606  
QY 2161 AGCTAGAGCTCTGATTCACATGCTGGTGAATGGGCAACCCAGCGGCAAGGATCAGA 2220  
Db 19605 AGCTAGAGCTCTGATTCACATGCTGGTGAATGGGCAACCCAGCGGCAAGGATCAGA 19546  
QY 2221 AACCAAGGCGCTGGGCAACTATGATGTCAAAAGAGATGGCTCAGAGTTGTGGGT 2280  
Db 19545 AACCAAGGCGCTGGGCAACTATGATGTCAAAAGAGATGGCTCAGAGTTGTGGGT 19486

QY 2281 AGAGGTCCAATCTGGGGGAGCCCAATTAAGTTCTGGGTGATTCAGATTAACCACTC 2340  
Db 19485 AGAGGTCCAATCTGGGGGAGCCCAATTAAGTTCTGGGTGATTCAGATTAACCACTC 19426  
QY 2341 ATGGCTTGTGTGCCATGAGTTAGGATGACAAAGTGAATGAAGTTGAAGTGGGAAACA 2400  
Db 19425 ATGGCTTGTGTGCCATGAGTTAGGATGACAAAGTGAATGAAGTTGAAGTGGGAAACA 19366  
QY 2401 GAAATACCAAGCTGTGTGTCAAGGCAAGCTGGAAGAGAGAGAAAGAAATGAATGGCA 2460  
Db 19365 GAAATACCAAGCTGTGTGTCAAGGCAAGCTGGAAGAGAGAGAAAGAAATGAATGGCA 19306  
QY 2461 CCAATGAGAGCAATTTGCAAGAACAGCTCCCTGGAGATCTTGTGGAGGCTCAGAGACTT 2520  
Db 19305 CCAATGAGAGCAATTTGCAAGAACAGCTCCCTGGAGATCTTGTGGAGGCTCAGAGACTT 19246  
QY 2521 GCTGGCAGAGAGATCTGAGCTTACCCCAATTAAGCTCTGGATATCTGACCAATCTAGACC 2580  
Db 19245 GCTGGCAGAGAGATCTGAGCTTACCCCAATTAAGCTCTGGATATCTGACCAATCTAGACC 19186  
QY 2581 AGCAATGTCACTGGCAAGAGGTTGCAAGTCTTGTGTTATTTTCTGTCAATAACTGGT 2640  
Db 19185 AGCAATGTCACTGGCAAGAGGTTGCAAGTCTTGTGTTATTTTCTGTCAATAACTGGT 19126  
QY 2641 AAGGCTTGGGTTCCAAATTTGCTGACAGCTGTTTAACTGGGAAATGGGCTTAGACTAT 2700  
Db 19125 AAGGCTTGGGTTCCAAATTTGCTGACAGCTGTTTAACTGGGAAATGGGCTTAGACTAT 19066  
QY 2701 GGTAGCTATGTCTCAGACAAAGGCGCTTATCTTCACTGCGCTTAAACCAAGCTGAGTT 2760  
Db 19065 GGTAGCTATGTCTCAGACAAAGGCGCTTATCTTCACTGCGCTTAAACCAAGCTGAGTT 19006  
QY 2761 GGAAGCTGAGCTTGTTCAGCTTCAAAAATTAAGCTGAGTTTCCAGCAAGGCGCTTAT 2820  
Db 19005 GGAAGCTGAGCTTGTTCAGCTTCAAAAATTAAGCTGAGTTTCCAGCAAGGCGCTTAT 18946  
QY 2821 CTGAGCTTCCGTTCTTACAGCTTCAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2880  
Db 18945 CTGAGCTTCCGTTCTTACAGCTTCAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 18886  
QY 2881 CTTTCAGTGAACATGAATATAGACTCAACCAATCTCTTGAATCTCTTGGGAAAGGCG 2940  
Db 18885 CTTTCAGTGAACATGAATATAGACTCAACCAATCTCTTGAATCTCTTGGGAAAGGCG 18826  
QY 2941 TCTGAGCTGAGACCAAGACTGTGCTCATGGCTCATGTGATCTGAGTCAAGCCCTCC 3000  
Db 18825 TCTGAGCTGAGACCAAGACTGTGCTCATGGCTCATGTGATCTGAGTCAAGCCCTCC 18766  
QY 3001 AACCTGTCAAGCAATTTGCTCCGTAAGACTTGTGATGGGTAGAGTATGATCAACACTCT 3060  
Db 18765 AACCTGTCAAGCAATTTGCTCCGTAAGACTTGTGATGGGTAGAGTATGATCAACACTCT 18706  
QY 3061 GACTGTCAACAAAGCTTGTGACTGGAGGCGCAAGCTATAGAGTGGCTCAGCTTAAAG 3120  
Db 18705 GACTGTCAACAAAGCTTGTGACTGGAGGCGCAAGCTATAGAGTGGCTCAGCTTAAAG 18646  
QY 3121 GCTGGAGCTGGGAGCAAGTGTCTCAAGTTAAGGCTTAACTAGAAATTTGACTGAGCTG 3180  
Db 18645 GCTGGAGCTGGGAGCAAGTGTCTCAAGTTAAGGCTTAACTAGAAATTTGACTGAGCTG 18586  
QY 3181 AGAACAAGGTTAAGGAGCCCAAGCAAGAGGTTGTGGGTCTAATCTTGAAGAGCACTTGA 3240  
Db 18585 AGAACAAGGTTAAGGAGCCCAAGCAAGAGGTTGTGGGTCTAATCTTGAAGAGCACTTGA 18526  
QY 3241 GCTTAACTTCTTAACTTCTAATGAGTGTGAGATGAGCTTCAACCGGAGTGGCTCTAG 3300  
Db 18525 GCTTAACTTCTTAACTTCTAATGAGTGTGAGATGAGCTTCAACCGGAGTGGCTCTAG 18466  
QY 3301 TCTTAAGGAGTGAAGCAAGAGCAAGATTTCAAGTCTGATCTTATCAAGTTCAATGACTTA 3360  
Db 18465 TCTTAAGGAGTGAAGCAAGAGCAAGATTTCAAGTCTGATCTTATCAAGTTCAATGACTTA 18406

QY 3361 CACTGGAGACACGCTGATCATGACGAGGAGGCTAGAGGTCTAAGATTTACCA 3420  
 DB 18405 CACTGGAGACACGCTGATCATGACGAGGAGGCTAGAGGTCTAAGATTTACCA 18346  
 QY 3421 CACTGGATATCCAGATTTACTTTCTTTGTTTCAAGAGGAGGAGTCTAAGTCTC 3480  
 DB 18345 CACTGGATATCCAGATTTACTTTCTTTGTTTCAAGAGGAGGAGTCTAAGTCTC 18286  
 QY 3481 CCCTTTCTCTCTCTCCCAAGTGTCTCTGACGAGGAATCTTACGTTGTCTCAGAGG 3540  
 DB 18285 CCCTTTCTCTCTCTCCCAAGTGTCTCTGACGAGGAATCTTACGTTGTCTCAGAGG 18226  
 QY 3541 ACTCCAGAAATGTTTGTTCAGTCACTTACGCTGCTATTAAGATATCTTAGAGT 3600  
 DB 18225 ACTCCAGAAATGTTTGTTCAGTCACTTACGCTGCTATTAAGATATCTTAGAGT 18166  
 QY 3601 GGGTAACTATCAGCAATAGAAATTTATGTTTCAATTTCTGAGAGGCTGAGAAATCCAG 3660  
 DB 18165 GGGTAACTATCAGCAATAGAAATTTATGTTTCAATTTCTGAGAGGCTGAGAAATCCAG 18106  
 QY 3661 ATCAAGGCTCAGAGGCTCAGTGTCTGCTGAGTGTGTTCTGCTTCAAGATGAGACC 3720  
 DB 18105 ATCAAGGCTCAGAGGCTCAGTGTCTGCTGAGTGTGTTCTTCTTCAAGATGAGACC 18046  
 QY 3721 TTTTGTCTGTGTCTCA 3737  
 DB 18045 TTTTGTCTGTGTCTCA 18029

RESULT 3  
 AC024232 148198 bp DNA linear HTG 07-JUN-2000  
 LOCUS Homo sapiens chromosome 11 clone RP11-223K12, WORKING DRAFT  
 DEFINITION  
 AC024232 148198 bp DNA linear HTG 07-JUN-2000  
 AC024232 9 unordered pieces.  
 AC024232.3 GI:8516169  
 HTG; HTG; PHASE1; HTG; DRAFT.  
 VERSION  
 HTG; HTG; PHASE1; HTG; DRAFT.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 1 (bases 1 to 148198)  
 Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 148198)  
 WATERSTON,R.H.  
 REFERENCE  
 Submitted (28-FEB-2000) Genome Sequencing Center, Washington  
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT  
 On Jun 14, 2000 this sequence version replaced gi:7715652.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H.NH022K12  
 ----- Summary Statistics -----  
 Sequencing vector: M13, 100%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.99019  
 Consensus quality: 142463 bases at least Q40  
 Consensus quality: 144172 bases at least Q30  
 Consensus quality: 145227 bases at least Q20  
 Insert size: 15100; agarose-fp  
 Insert size: 147398; sum-of-contigs  
 Quality coverage: 4.53 in Q20 bases; agarose-fp  
 Quality coverage: 4.67 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2755: contig of 2755 bp in length  
 2756 2855: gap of unknown length  
 2856 12253: contig of 9398 bp in length  
 12254 12353: gap of unknown length  
 12354 22567: contig of 10214 bp in length  
 22568 22667: gap of unknown length  
 22668 37602: contig of 14935 bp in length  
 37603 37702: gap of unknown length  
 37703 51053: contig of 13351 bp in length  
 51054 51153: gap of unknown length  
 51154 66793: contig of 15540 bp in length  
 66794 66793: gap of unknown length  
 66794 85880: contig of 19087 bp in length  
 85881 85980: gap of unknown length  
 85981 115955: contig of 29975 bp in length  
 115956 116055: gap of unknown length  
 116056 148198: contig of 32143 bp in length.

## FEATURES

## source

1. 148198  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /clone="RP11-223K12"

1. 2755  
 /note="assembly\_name:Contig4"

2856. 12253  
 /note="assembly\_name:Contig5"

12354. 22567  
 /note="assembly\_name:Contig6"

22668. 37602  
 /note="assembly\_name:Contig7"

37703. 51053  
 /note="assembly\_name:Contig8"

51154. 66693  
 /note="assembly\_name:Contig9"

clone end: T7  
 vector\_side: right"

66794. 85880  
 /note="assembly\_name:Contig10"

85981. 115955  
 /note="assembly\_name:Contig11"

clone end: SP6  
 vector\_side: right"

116056. 148198  
 /note="assembly\_name:Contig12"

BASE COUNT 38577 a 34635 c 33776 g 40404 t 806 others

ORIGIN

Query Match 65.5%; Score 2449; DB 2; Length 148198;  
 Best Local Similarity 98.7%; Pred. No. 0;  
 Matches 2490; Conservative 0; Mismatches 30; Indels 3; Gaps 2;

QY 1 CTCGAAGAGAGGCTCAGTCAAGCTGACGAGACAGATTGGAAGTCAATGATATAGAT 60  
 DB 20008 CTCGAAGAGAGGCTCAGTCAAGCTGACGAGACAGATTGGAAGTCAATGATATAGAT 20067  
 QY 61 GGCATTTAAAGCCCTGAGCTAGTGAATTAACCAAGAGTGAAGTGAAGAGAGA 120  
 DB 20068 GGCATTTAAAGCCCTGAGCTAGTGAATTAACCAAGAGTGAAGTGAAGAGAGA 20127  
 QY 121 GAAGAGGCCCAAGTGAAGTGAATTTAATTAAGATATGATGATGAAGAAAGATAGTC 180  
 DB 20128 GAAGAGGCCCAAGTGAAGTGAATTTAATTAAGATATGATGATGAAGAAAGATAGTC 20187

Db	4632	TTTTTTATGTCGAGATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT	4691
Qy	1692	TGACTTATGATGTCCTGGACCCATTGTCAGTGTGAGAGTCAACGTGAGCCTCAGCAGTG	1751
Db	4692	TGATTTATGATGATTTGGATTTAATGTTAGTGTGAGAGTATTAAGTTGACGTTAGTACTG	4751
Qy	1752	TAGCCAGTTCCTGCTGTAATAATTCGTAAGAGGGGTTGGGGGGCAGCTCCGGGAAAAG	1811
Db	4752	TAGTTTATGTAATGTTTGAATAATGTTGAAGAGGGGTTGGGGGGTATGTTGTGGGAAAAG	4811
Qy	1812	GAGCTTGATTCAGATTTCTGTGTCCAGACCCCTGACCTTATTTGTCAGTATGAATCAGCC	1871
Db	4812	GAGTTTGGATTTAGATTTTGTGTTAGATTTTGAATTTTATTTATTTGTAATGTAATTAAGTT	4871
Qy	1872	AATATGCGCTTAGTCTCGGAGACAGACATTTCCAGTAGAGTTGAGAGTGGGGGTGTG	1931
Db	4872	AATATGCGCTTAGTCTCGGAGATAGTAATTTTATTTAGTAGAGTTGAGAGTGGGGGTGTG	4931
Qy	1932	CTGCTGCCAATCTATATATAGGAGTTCACATGCTACCCAGACGCTCTCTGTGACCTGTG	1991
Db	4932	TTGTTGTTAAATTTATATATAGGAGATTAATTTAGTTAATTAAGTTGTTTGTGTTTGTG	4991
Qy	1992	CAGCTCAGCATGAGTGGGATCTGGAGACACCCGTTGACACTGGGGTGTGAGCCTATGC	2051
Db	4992	TAGTTTATGATGATTTAGGATTTGGGATATTCGTTGATTTGGGGTGTGAGATTTATGT	5051
Qy	2052	TGCTCTCGCCATTCGCCACCCCTCTTCTCCGTAGATTAAGCTGGAGCTGAAGCGAAG	2111
Db	5052	TGCTTTTGTGTTATGTTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG	5111
Qy	2112	GATTGAGTTCTGGGCTAGGGTATAGGTAGGGCCAGTTTATAGGCTCGGTCAAATTTGGGG	2171
Db	5112	GATTGAGTTTGGGTTAGGGTATAGGTAGGGCTTAAATTTAGGTTTGGGTTAAATTTGGGG	5171
Qy	2172	TCAGGGGCTATGGGAAAGAGATCGGCTCCCATATGATCAAGATATCTATTTGTTCTCCT	2231
Db	5172	TTAGGGGCTATGGGAAAGAGATCGGCTTAAATGATTAAGATTAATTTATTTGTTTGT	5231
Qy	2232	AGG 2234	
Db	5232	AGG 5234	
RESULT 6			
AX345847/C			
LOCUS	AX345847	5234 bp	DNA
DEFINITION	Sequence 918 from Patent WO0200928.		linear
ACCESSION	AX345847		
VERSION	AX345847.1	GI:18493733	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.		
JOURNAL	Diagnosis of diseases associated with the immune system		
	Patent: WO 0200928-A 918 03-JAN-2002;		
FEATURES			
Source			
	1..5234		
	location/Qualifiers		
	/organism="synthetic construct"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32630"		
	/note="chemically treated genomic DNA (Homo sapiens)"		
BASE COUNT	1303 a	40 c	1091 g
ORIGIN			
	12.8%; Score 1281.6; DB 6; Length 5234;		
	Best Local Similarity 73.4%; Pred. No. 0;		
	Matches 1638; Conservative 0; Mismatches 594; Indels 0; Gaps 0;		
Qy	1	TCCTCTCCAGGAGCCAGCAAAATCTGTAGATTCAAGACGGGTTCTGACAGCTG	60

Db	2234	TCCCTCTCCCAACCAACCAACCAAAATCTATAAATTCAAACAAATTTTAACTA	2175
Oy	61	AAGCAAGTGTGAGAAATTCCTGTAGAGATCATGGGCTCTCAGAGGAGATA	120
Db	2174	AAACAATTAATTAATAAAATTCCTAATTAATAAAATCAATAAATCTCAAAAAAATA	2115
Oy	121	TAAGTTTCAGAGCTCAGAGGAAAGAAAAAGTGAAGGGAGCTTAAATTAAGTCTC	180
Db	2114	TAAATTTCAAAAATAAAAAAATAAAAAAATAAAAAAATCTTAATAATTAATCTC	2055
Oy	181	CCATTGCCAACACCCAGAAAGAAGATGCGCCCTGCATGGGAGGAAGTGAATGAGA	240
Db	2054	CCATTACCAACACCCAAAAAATAAAACATACCCTCAATAAAAAATAAATATTAATA	1995
Oy	241	CATTGGCTGAGCAGCATGGCATTTGCCAGGCTGCCAAGACTCAGAGTCCAGCTT	300
Db	1994	CATTAACTATTAACAACATTAATTAACCAAACTATTAACCAAAATCTCAAAAATTC	1935
Oy	301	GCCCACTGACTATAGAGAGGAAATGATGTTCAAGCACTTTTCATTCGTAAGTCAGA	360
Db	1934	ACCACATTAACCTTAATAAAAAAATAATTAATTAATTAATTAATTAATTAATTA	1875
Oy	361	GAGGACATGAGCGCTGATGGCAGAGCGCTGGTGCATGTTGTCAGAGGTTCCGAGT	420
Db	1874	AAAAACATTAACCTTAATAAAAAAATAATTAATTAATTAATTAATTAATTAATTA	1815
Oy	421	TGTTGTTTCTGTGTAAGAAATCTTCGACAGATGAAAGGATCTGAGACTTTGGTA	480
Db	1814	TATATTTTCCATTAATAAAAAAATTCGCAAAATTAATAAAAAAATCTTAATTAATA	1755
Oy	481	AGATTATATATGAGACTGTCAAGGGCTTGGAGCATCTGAGGAGATCAGGCGCTTCA	540
Db	1754	AAATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1695
Oy	541	GCTTGGCTGAGGAGCGGGTCCGTGAACTTAATCCGCGCATAGCTGAGTCTGCCA	600
Db	1694	ACCTTAATTAATAAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1635
Oy	601	TAAATTTTCTGACTCACTAGCAAAATCTCAACAAGAAATGGGGAGCTTTGGGAGTG	660
Db	1634	TAAATTTTCTGACTCACTAGCAAAATCTCAACAAGAAATGGGGAGCTTTGGGAGTG	1575
Oy	661	GCCAGAGAGTACTGAGGATGACGGTGAATCCAGAGAGAGATGATGTGGGCGGAG	720
Db	1574	ACCAAAAAATTAATAAAAAATTAATAAAAAATTAATAAAAAATTAATAAAAAATTA	1515
Oy	721	ACACTGAGAGAGAAACAGACTGTCAATTAAGGGGCTCTGTGACTCTTAATTAATTA	780
Db	1514	ACACTTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1455
Oy	781	ATGCTTACTACATTAACCTACCCCCAATTCCTAATTTCTCTACCTTGAAGGGGGAA	840
Db	1454	ATACTTACTACATTAACCTACCCCCAATTCCTAATTTCTCTACCTTGAAGGGGGAA	1395
Oy	841	ATTGTGAGAAATTTGGTCAACATGACACACTACTCACTACTGTAATGCAATTTTG	900
Db	1394	ATTATCAAAAATTAATACTAACAACCTAACAACAATACTTAATAATACTTAATTTTA	1335
Oy	901	CATTTTTCATTAACAATATTTCTGGAACAATCTTAATAGCCAGGCTATTTTGA	960
Db	1334	CATTTTTCATTAACAATATTTCTGGAACAATCTTAATAGCCAGGCTATTTTGA	1275
Oy	961	GAGTCAGGAGATTAATAATGTAACAAGCAGGCAAAACAAGCAAAAGCAACAACA	1020
Db	1274	AAATCAAAAATTAATAATTAATAACAACAACAACAACAACAACAACAACAACA	1215
Oy	1021	TCACCAATTAAGTACAGATGAAGAATTTCAAGTTTATAGTAATTAATAACAAGC	1080
Db	1214	TCACCAATTAATTAACAATTAATAAAAAATTTCAATTTTATTAATTAATAACAAGC	1155
Oy	1081	AAGGCTGAAATGGCTGATAGCGGCTCAGAAAGGCTTCATTGAGAGGTAGCATTT	1140

```

Db      1154 AAAATCTAAATACTAAATAATATATCAAAAAAACTTCAATTAATAATAATACATT 1095
QY      1141 AACGAGAGTCAGTGAATAATATGTGAATTCAGTTAGATGTTATTTGTTGGGT 1200
Db      1094 AAACAAATCAACCTAAATATTTATTAATTTCAATTCATTAATTTATTTAAAT 1035
QY      1201 GGTAAATAAAGCTTTTCCCAAGGTGAAAATACCAAGAAAGACTAATTAAGTAG 1260
Db      1034 AATTAATATAAATCTTTTCCCCCAATAAATACTACCAAAAAAACTAATTAATA 975
QY      1261 TGGTGGTCTCTGAGAGAGACACCTCTGTTTGTGCTCATTAATCTCAACCTTC 1320
Db      974 TAATAATACCTCTAAAAAAAACACTCTTAATTTCTACCTCATTAATCAACCTTC 915
QY      1321 ACTTCAGGACCTTTTGGCAAGCCCTTTCAGTCAGGAGGAGGAGGAGGAGGAGGAG 1380
Db      914 ACTTCAGGACCTTTTGGCAAGCCCTTTCAGTCAGGAGGAGGAGGAGGAGGAGGAG 855
QY      1381 GGGCTTGAATTTGACACAGTGAACATTAATTTGCTCCAGATCTAGCCCAAGGAT 1440
Db      854 AAATTAATCAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 795
QY      1441 AAAGCTGAAGAGCTTGGGATGCCCCGAGAAAGCCCTGATGAGCTTGAGAAAGCTG 1500
Db      794 AAACCTAAAAAACTTAAACATACCCCAAAAAAACCCTAATTAATTAATTAATTAAT 735
QY      1501 TTCTGAGATTTCTAAGTAAGTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
Db      734 TTCTGAGATTTCTAAGTAAGTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 675
QY      1561 TGTGTCTGTAGTGTGCTTGAAGCAGGAGCTTGAACAGTACAGACATTAATTTCCCT 1620
Db      674 TACTATCTTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 615
QY      1621 CCTTTCTCACTCCCTCCCAATTTGAGATTAATCAATCAAAAGTATCTCTCACTCT 1680
Db      614 CCTTTCTCACTCCCTCCCAATTTGAGATTAATCAATCAAAAGTATCTCTCACTCT 555
QY      1681 ACTCACTCCCTGACTTATGATGCTGAGACCAATTCAGTGTGAGAGTACAGCTGA 1740
Db      554 ACTCACTCCCTGACTTATGATGCTGAGACCAATTCAGTGTGAGAGTACAGCTGA 495
QY      1741 CGTGAGAGTGAAGCCAGTACTGCTTGAATAATGCTGAAGGAGGAGTGGGGGAGCTG 1800
Db      494 CGTGAGAGTGAAGCCAGTACTGCTTGAATAATGCTGAAGGAGGAGTGGGGGAGCTG 435
QY      1801 CCGGAAAAAAGAGCTTGTGATTCAGATTTCTGTCAGACCTGACCTTAATTTGCACTGA 1860
Db      434 CCGGAAAAAAGAGCTTGTGATTCAGATTTCTGTCAGACCTGACCTTAATTTGCACTGA 375
QY      1861 TGTATCAGCATATTTGCTTGTAGTCTGAGAGAGACAGACATTCAGTGTGAGAGG 1920
Db      374 TATATCAACCATATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 315
QY      1921 TGGGGGTGTGCTGCTGCACTCTAATATGAGAGTTCAGTGTGACCCAGAGCTGTCC 1980
Db      314 TAAAAATAATATCTACTACCACTCTAATATGAAAAATTTCAATCAATCAATCAATCAAT 255
QY      1981 TGTGGCTCTGACGCTCAGATGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 2040
Db      254 TATATCTCTGACACCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 195
QY      2041 GAGAGCTATGTGTGCTGAGCAATTCAGCCCTCTTCTGCTGAGTGAAGTGGAGAC 2100
Db      194 AAAACCTATATCTATCTTACCAATTCACCCCTCTTCTGCTGAGTGAAGTGGAGAC 135
QY      2101 TAGAAGCGAAGATTGATTCTGAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 2160
Db      134 TAAAAAGCAAAATTAATTTCTAATCAAAATAAATAAATAAATAAATAAATAAATAAATAA 75
QY      2161 CAATTTTGGGGTCAAGGGGCTAATGGAAAGGAGATCGCTCCCAATGATCAAGATATCTATT 2220
Db      74 CAATTTTAAATCAAAAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 15

```

```

QY      2221 TTGTTCTCCCTA 2232
Db      14 TTATTTCTCCCTA 3

RESULT 7
HUMHXMA01
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
source
location/Qualifiers
1..731
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="11p15.5-p15.4"
178..288
/partial
/gene="HPX"
/note="G00-120-054"
/number=1
206..274
/gene="HPX"
/note="G00-120-054"
/number=1
289..438
/gene="HPX"
/note="G00-120-054"
/number=1
439..497
/gene="HPX"
/note="G00-120-054"
/number=1
498..617
/gene="HPX"
/note="G00-120-054"
/number=2
618..689
/gene="HPX"
/note="G00-120-054"
/number=3
690..731
/gene="HPX"
/note="G00-120-054"
/number=3

Intron
BASE COUNT 148 a 167 c 204 g 212 t
ORIGIN
Query Match 5 8%; Score 681.4; DB 9; Length 731;
Best Local Similarity 98.1%; Pred. No. 5e-170; Indels 3; Gaps 2;
Matches 711; Conservative 0; Mismatches 11;

QY      1796 AGCTGCGGGAAGAGAGCTTGTGATTCAGATTTCTGTCCAGACCTGACCTTATTTGC 1855
Db      1 AGCTGCGGGAAGAGAGCTTGTGATTCAGATTTCTGTCCAGACCTGACCTTATTTGC 60
QY      1856 AGTATGTAATCAAGCAATATTTGGCTTAAGTCTGTGAGAGACAGACATTTCCAGTAGATT 1915

```



JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
REFERENCE Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 249769)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
COMMENT On Aug 25, 2002 this sequence version replaced gt:22138702.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: M\_BA0042113  
----- Summary Statistics -----  
Sequencing vector: M13, 0%  
Sequencing vector: plasmid, 100%  
Chemistry: Dye-terminator Big Dye, 100% of reads  
Chemistry: Dye-terminator Big Dye, 100% of reads  
Assembly program: Phrap, version 0.99019  
Consensus quality: 240797 bases at least Q40  
Consensus quality: 242108 bases at least Q30  
Consensus quality: 242698 bases at least Q20  
Insert size: 204000; agarose-fp  
Insert size: 248995; sum-of-contigs  
Quality coverage: 10.08 in Q20 bases; agarose-fp  
Quality coverage: 8.91 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1085: contig of 1085 bp in length  
\* 1086 1185: gap of unknown length  
\* 1186 3097: contig of 1912 bp in length  
\* 3098 3197: gap of unknown length  
\* 3198 5110: contig of 1913 bp in length  
\* 5111 5210: gap of unknown length  
\* 5211 7669: contig of 2459 bp in length  
\* 7670 7770: gap of unknown length  
\* 7771 13621: contig of 5852 bp in length  
\* 13622 13721: gap of unknown length  
\* 13722 19656: contig of 5935 bp in length  
\* 19657 31061: gap of unknown length  
\* 31062 31161: contig of 11305 bp in length  
\* 31162 40106: contig of 8945 bp in length  
\* 40107 40206: gap of unknown length  
\* 40207 50613: contig of 10407 bp in length  
\* 50614 50713: gap of unknown length  
\* 50714 61953: contig of 11239 bp in length  
\* 61954 62052: gap of unknown length  
\* 62053 74978: contig of 12926 bp in length  
\* 74979 75078: gap of unknown length  
\* 75079 92714: contig of 17636 bp in length  
\* 92715 92814: gap of unknown length  
\* 92815 124084: contig of 31270 bp in length  
\* 124085 124184: gap of unknown length  
\* 124185 151377: contig of 27193 bp in length  
\* 151378 151478: gap of unknown length  
\* 151479 178112: contig of 26635 bp in length  
\* 178113 178212: gap of unknown length  
\* 178213 208382: contig of 30170 bp in length  
\* 208383 208482: gap of unknown length  
\* 208483 249448: contig of 40966 bp in length  
\* 249449 249548: gap of unknown length

FEATURES  
Source  
1. 249769  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP23-42113"  
1. 1085  
/note="assembly\_name:Contig33"  
1186. 3097  
/note="assembly\_name:Contig84"  
3198. 5110  
/note="assembly\_name:Contig85"  
5211. 7669  
/note="assembly\_name:Contig86"  
7770. 13621  
/note="assembly\_name:Contig87"  
13722. 19656  
/note="assembly\_name:Contig88"  
19757. 31061  
/note="assembly\_name:Contig89"  
31162. 40106  
/note="assembly\_name:Contig90"  
40207. 50613  
/note="assembly\_name:Contig91"  
50714. 61952  
/note="assembly\_name:Contig92"  
62053. 74978  
/note="assembly\_name:Contig93"  
75079. 92714  
/note="assembly\_name:Contig94"  
92815. 124084  
/note="assembly\_name:Contig95"  
124185. 151377  
/note="assembly\_name:Contig96"  
151478. 178112  
/note="assembly\_name:Contig97"  
178213. 208382  
/note="assembly\_name:Contig98"  
208483. 249448  
/note="assembly\_name:Contig99"  
249549. 249622  
/note="assembly\_name:Contig24"  
249723. 249769  
/note="assembly\_name:Contig56"  
/note="assembly\_name:Contig56"  
BASE COUNT 67104 a 56647 c 55906 g 68301 t 1811 others  
ORIGIN  
Query Match 4.8%; Score 479.6; DB 2; Length 249769;  
Best Local Similarity 55.1%; Pred. No. 4.1e-116;  
Matches 2135; Conservative 0; Mismatches 1309; Indels 430; Gaps 44;  
QY 3360 TTCTTGGGAGATCATCTAGCTTCATATACCTTCCTCAGTATCAACATAC 3419  
Db 125282 TCGTTGGGAGACATCAAGACTTTTATATACCTCCCTGGATATTCAGCCACTGC 125341  
QY 3420 CTGGCTTGAGGAATATATCTATCTGAGACATGATCTAATCTTCTGAGCTCC 3479  
Db 125342 CATGGCTTGAGGAGC--TCGTGCTCTAATATACAGATCTGATCACTCCCTGAGACTTT 125399  
QY 3480 AGACATCTTTCATATGTATCTTA-----CAGGACCTAAATTC 3523  
Db 125400 GGAATACATTTTTCATTTTGTGAAATCTGAGCTTTCCCGAGATACCTCAATATGC 125459  
QY 3524 AGCATCCCAACTAGCTTTCATCTTATCAACCAACCTTCTCCTGCTGCTTC 3583  
Db 125460 AGCATCTCTTAATATATTTACATCTTTTAAAGGACCTTCTCCTGCTGCTTC 125518  
QY 3584 CTGTTTCACTTAATGACCCCAAAATGCTGATTAATCAAAACCAAGTGACACAGGGT 3643



Db 125519 CTGATTCACTAGAG-GGTACCATTGCTGTAAGGTTACCAATGTTCCAGTGTAC----- 125571  
 QY 3644 CTCATGATCGGCGCTTGGTTATCTTCTGAGTTTATCTCCCTCCGCAACTTCAGTC 3703  
 Db 125572 -----TGGCCCTGAGGAGACTATCTTCTGCTGTTT-----CTTCCATGTTTACAG 125616  
 QY 3704 TGTGACAGCCATACGATCTACATGAGTTGAGACACACTGTTCCCTGATTTGGGCTC 3763  
 Db 125617 TGTACAGTCGCAATAGTTTATCCAAAG-TCAATCCAAATCTTATACATGTTACAGTCC 125675  
 QY 3764 TGCATGCTCTCCCTGCTGCTGTAACACCTTTCTCATCTTGTCAACCTGGAATAATCTCT 3823  
 Db 125676 TCTGTCTCTCTGTC-----CTTGTGAAAATCTCT 125703  
 QY 3824 GCTGATTTTTCAGCTCTTGGGCCCAATGCTTCTCTTGTGTGAAAACCTTCCACAAT 3883  
 Db 125704 CACTCATTTTTCATTTCTTGGCCATATATCTTCTCCACTGTGAAAATTTCTGGGCTCT 125763  
 QY 3884 CTCTAGGACACTTACGACACTCTGTCTATATCTCAGTGCATCTTACACTACACTTGG 3943  
 Db 125764 TTCCAGACATATTAATGACATTTCCCCCAATTTCTCAATGCTTGTGATTT---CTTTT 125820  
 QY 3944 GATGTTGATGCTGTAAGATGACAGATGCTCTTCTGTTTGTACAGTGAATCTCTGAA 4003  
 Db 125821 TATCTTGTGCTGAACGGCTGGGATGATGATGTTGTTTCAAGGATCTT----- 125874  
 QY 4004 GTGAAAAGACAGATCTTGTATCTCTCAGTGCCTCTCACAATGCTCTGATATGTTAT 4063  
 Db 125875 -----GTCAATCCGAAATGTTCAAGCTCTTGTGTGTGAGCTGTGACTGTATGTA 125925  
 QY 4064 TTCAAGTACTGTTCTTCTGATGATGAATGAATGAATGAATGAATGAATGAATGAATGA 4123  
 Db 125926 TTCAAGCGCTGTTCTTGAAGAAATCCGTAATGATGAGCAATGATATGATGAGG 125985  
 QY 4124 GAAATACATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4183  
 Db 125986 CAA-----GTGATTAAGAGATTAATCAATGATGATGATGATGATGATGATGATG 126034  
 QY 4184 GTCTTCAGATTAATATGTCACAGAGATCTCTGG--GTACAGCTGTAAGAGCAATGTGT 4241  
 Db 126035 GAGCTGTATTAATGATGTCACAGGAGATCTGTGTATATCATGTCAGAGGTGATGTCAT 126094  
 QY 4242 TATAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4301  
 Db 126095 TCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126147  
 QY 4302 CATTTGTTCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4361  
 Db 126148 TGTGTATCTCACTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 126207  
 QY 4362 CAGTATATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4421  
 Db 126208 CATATAGCAAAAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 126267  
 QY 4422 AG-----AGAGTACTT 4432  
 Db 126268 TGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 126327  
 QY 4433 TTGATGACAGGAGGAGAAAGTCTGGTATATCCTCTGTAAGAAAGAGAAAGATATCC 4492  
 Db 126328 TTGCTGACAGGAGAAAGTCTGGTATATCCTCTGTAAGAAAGAGAAAGATATCC 126387  
 QY 4493 AAAGTGTCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4552  
 Db 126388 AAAGTGTCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126447  
 QY 4553 TCACCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4612  
 Db 126448 CCACCGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126507  
 QY 4613 ATCCAAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4672  
 Db 126508 AACCAAGCAC-----ACAGATGAGATCATGTGTC 126536

QY 4673 TAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4732  
 Db 126537 TATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126555  
 QY 4733 CGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4792  
 Db 126596 GTTACAGGCTAGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 126654  
 QY 4793 GTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4852  
 Db 126655 -----CTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126697  
 QY 4853 TCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4912  
 Db 126698 TTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126752  
 QY 4913 CTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4972  
 Db 126753 -----ACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126796  
 QY 4973 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5032  
 Db 126797 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126827  
 QY 5033 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5092  
 Db 126828 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126882  
 QY 5093 AGACTGTCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5152  
 Db 126883 CAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126936  
 QY 5153 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5212  
 Db 126937 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126993  
 QY 5213 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5272  
 Db 126994 ATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127052  
 QY 5273 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5332  
 Db 127053 GAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127112  
 QY 5333 GAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5392  
 Db 127113 GAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127172  
 QY 5393 TGGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5452  
 Db 127173 TTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127232  
 QY 5453 AGGTGCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5512  
 Db 127233 AGGTGCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127292  
 QY 5513 GTGAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5572  
 Db 127293 GTGAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127352  
 QY 5573 ATTCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5632  
 Db 127353 GTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127411  
 QY 5633 GCC-CCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5691  
 Db 127412 CCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127470  
 QY 5692 AACCCAGCATCTTCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5751  
 Db 127471 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127530

QY	5752	CCCTACCCCAATTTCTCTCTCCAGGC--CATGACACAGGAAATGGACCTGGCCATGGGAAC	5809
Db	127531	CCATBACCTGTTTCTCTCCAGGCCATGGTAGACCCAGAAATGAACCTGTACATGGAAAT	127550
QY	5810	AGTACCCACCATGAGCCCTGAGTATATGACGCTGTAGGCCACATCTAGTCTGTGTGCACTG	5869
Db	127591	AGCACCCA-----TCTATGTGATTCACGTTTAGGCCAGATCTGGCCGTACGTGACCTG	127644
QY	5870	ACGTCTGACACCAATGATGTGCCACTATGCTTCAATGTGTGAGAGATGCCCCCACTCC	5929
Db	127645	CTGTGTGACCATGCGGGGTGCCACATATGCTTCACTGTGTGAGAGGCGTGCATCTCAGATCCC	127704
QY	5930	CAATGTGCTCTCACATCTCTTTTACTGTGTATCCATCCCTGACCATTTCTCCATTTGT	5989
Db	127705	CATGTCTACCCCCCCCCACCCCGTCTCTTTTACTATATGTCTGTCTTTTACAGATAGCTT	127764
QY	5990	CATCACTGTGTCACTTATTTTGTGCCCTGTGCCCATCTTCTGTGACGCTTCTGCAT	6049
Db	127765	CTCCACTTTTACACACACATCTCATATTTGGCCCTTTGTATCTCTCTCCATACATACAT	127824
QY	6050	CCCTCATCT-CTGAGGCATATTTCTCATCTGTGTGTGTGTCAAGGCCAAG---CCCTTAA	6104
Db	127825	CCCTCATCTCTCCAGGTTTCATGTGCACCACTTATCTCATATCAGAGCCCTTCTTCA	127884
QY	6105	CTTATCTTACCGTGTACATCTACCTCCATGGCGTGTGCCCTGTGACCTCTGTGGG	6164
Db	127885	TGCTTTTATCCCTGTCTGTCTTCAAGTGTATCTACACCCCTGTGTGTATTTCTGTGGGA	127944
QY	6165	CCCTATGACTCCCTGTGTGTCTC---CTTGCTCAATGCCCTGTGAGCCCTGTGGCTCTCC	6221
Db	127945	CCCTCTTATCCCATATATATCTTCTCATAGTTTCTCTCTGTGTGTGTGTGTGTGTGTGA	128004
QY	6222	CTTGTCTCCCTGGAAGCTCTATATGTGTCTGTGTACCTCTGTCCCTTGTGTCTTGCAAT	6281
Db	128005	CTTGTCTCTGTGAGCTGTGTGTGTGCATG--CCGTGTGCATCCC-TGTGTGTGTGGCAT	128061
QY	6282	CTTTCTGAGTCTCTGTGGCTC-----	6301
Db	128062	CTCAGTGAAGCTCCGT	128121
QY	6302	-----CCCTGATTTATCTCA	6318
Db	128122	GTGTGATTTGTGTAAGTACATAGCCCTTTTGTAGTATATGTGATTCATTAATTAATTTCA	128181
QY	6319	GAACTCCATCTTGT	6378
Db	128182	GGAACCCCATTTTTTCAAGGTACTTATGTTCTTATCATCCAGGTCTGTGGCTGTGTGTGTGT	128241
QY	6379	CTGGGGAGAGATGTGTCTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	6438
Db	128242	TTGGGGAGTGAAGTGTCTCTCAATGTCTTAAAGAAACAAAGAAAGAAAC--TGAAATATTCTTGAAC	128299
QY	6439	TTTATGAGCTGGCTTCACTAGAGCTGTGAGGTTTCTCTGTAGTACCAATGAGATATAC	6498
Db	128300	CTTACGATGTCCCTTCA-----AAGAGATCTCAAGAAAGAA	128336
QY	6499	CCCTCCCTTGAACCCGTGCATCATTTGTGTAAGCCATCTTAAATTAATCTTAGGCGCTGT-T	6557
Db	128337	CCCAACCTTGAAGCTTTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	128396
QY	6558	CTGAGTTTATGTACGAGCAGTAAATATGTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT	6617
Db	128397	AATCTCTTAGGCTGTGAGGTGTACCACTCTCTACCTGTGAGCACTCCCAAAATGGGGAG	128456
QY	6618	TGAGCAACCATCATCATCAACCTGTAGATATATGTATATTAATTAGGACAGTGTGAAGATA	6677
Db	128457	GGGACATATCATATCATCTTACCTTACCTGTAGTATATGTCA-----ACAAATGTGTCTTAAATG	128511
QY	6678	TAAATATGTGAAATATATTTTTTTCACACAAATTTTTTGTGCTCTGTGACCTTGTGACAA	6737
Db	128512	GCAAGGGGACAGAAATAT-----AGAAATATCTTTTATGTGTGTGTGTGTGTGTGTGTGTGT	128564
QY	6738	TTTGTACCATTTATGTATCAATGTTCTGTGTGAAAAATATCATCAACATGTAGAGCAAT	6797

[illegible]

RESULT 10	
CNS01DM6/c	
LOCUS	CNS01DM6 202495 bp DNA linear PRI 28-APR-2001
DEFINITION	Human chromosome 14 DNA sequence BAC R-596D21 of library RPc1-11
ACCESSION	from chromosome 14 of Homo sapiens (Human), complete sequence.
VERSION	AL136418
KEYWORDS	AL136418.4 GI:13897283
SOURCE	HTG.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Mumukshu, Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Metazoa; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 202495)
AUTHORS	Hellis, R., Pettit, J.L., Vico, V., Dasilva, C., Robert, C., Winkler, P., Broctier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F., Leyva, M., Eckenberg, R., Brula, T., deBerardinis, V., Cruaud, C., Gysay, G., Saurin, W. and Weissbach, J.
TITLE	Sequencing of the human chromosome 14
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 202495)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (26-APR-2001) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT	- Web : www.genoscope.cns.fr On Apr 30, 2001 this sequence version replaced gi:12001732. ----- Genome Center Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: SeqRef@genoscope.cns.fr ----- The following BAC sequence is oriented from the T7 to the SP6 end. Upstream BAC (overlapping the T7 end) : R-176H8 (AC=AL139353) Downstream BAC (overlapping the SP6 end) : C-2313013 ----- Summary Statistics Assembly program: Phrap, version 2.0 Quality coverage: 7.09x in Q20 bases; sum-of-contigs ----- Overall quality chart : Range : bases 0 : 1 - 9 : 1 10 - 19 : 9 20 - 29 : 112 30 - 39 : 2205 40 - 49 : 7368 50 - 59 : 7368



60 - 69 : 11469  
70 - 79 : 28289  
80 - 89 : 66625  
90 - 99 : 86417  
-----  
Percentage of bases with a quality value >= 40 : 99 %  
Location/Qualifiers  
1. 202495  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-596D21"  
/clone\_1ib="R-PCI-11"  
18107..18281  
/note="matching EMBL:R22776  
RHD:RH6015  
dbSTS:STS45945  
Identified using the e-PCR software (G. Schuler)"  
STS  
23355..23600  
/note="matching EMBL:AA521333  
RHD:RH98883  
dbSTS:STS68632  
Identified using the e-PCR software (G. Schuler)"  
STS  
169462..169599  
/note="matching EMBL:R69806  
RHD:RH53627  
dbSTS:STS29238  
Identified using the e-PCR software (G. Schuler)"  
BASE COUNT 57490 a 41081 c 42091 g 61833 t  
ORIGIN  
Query Match 4.8%; Score 479.4; DB 9; Length 202495;  
Best Local Similarity 54.9%; Pred. No. 4.6e-116;  
Matches 1436; Conservative 0; Mismatches 1011; Indels 170; Gaps 18;  
QY 7181 ATATTAATAATACCGAGCGCGCGGTGAGTCAAGCCCTGTAATCTGACACTTTGGGA 7240  
Db 97791 AAAAAAAAAAAGAAAGCCGGCAATGTTGCTAATGCTGTATATCCAGACACTTTGGGA 97732  
QY 7241 GCGTGAAGCAGTGAATGCTTGAAGCGGCTTGAAGCCTGAGAGTTTGAACAGCCTGAGC 7300  
Db 97731 GCGCGAGCGAGTGAATCAAC-----TGAGTCAAGAGTTTCAAGACCACTTGGTC 97681  
QY 7301 AACATGCGAAACCTCATCTCTACAAAAATATTAATAATTAATGTCGGGTGTAATGCGT 7360  
Db 97680 AACATGCGAAACCTCATCTCTACAAAAATATTAATAATTAATGTCGGGTGTAATGCGT 97621  
QY 7361 TCCTATAGTCCATCTACTTCAAGAGCGTGAAGCCGGA-----AGGTC 7402  
Db 97620 GCGTGTATATCCAGCTACTCGGAGGCTGAGGAGGAGAAATCACTTGAACTGAGAGTG 97561  
QY 7403 GAGGCTTCAGTGAAGCGGTGATCGTGTACTGCACTCCAGCCTGGTGAAGAGTGAAGC 7462  
Db 97560 GAGGCTTCAGTGAAGCGGTGATCGTGTACTGCACTCCAGCCTGGTGAAGAGTGAAGC 97501  
QY 7463 ATGTCTCAAAAAAACAACAAACAACAAACAACAAACAACAAACAACAAACAACAAACAAC 7522  
Db 97500 CCGCTTCACAAAAAACAACAAACAACAAACAACAAACAACAAACAACAAACAACAAACAAC 97441  
QY 7523 AT 7582  
Db 97440 CCGTTTCAAT 97381  
QY 7583 ATGACCCAGCTACAGATGGG-----CAGTCCCTCCCTCTCACTGGTAA 7628  
Db 97380 GTCTGAGGCTTACGTTGTGTGTCAGTGTAGACATTAAGACATCGGTATATATATAT 97321  
QY 7629 ATTTTCTTTCTCTGACTCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7688  
Db 97320 AT 97261  
QY 7689 TCACTGTGACCCAGGCTGAGTGAAGTGGCGCAATCTTGTTCACCTGCAACTCTGCC 7748

Db 97260 TTGCTCTGTCAACTAGGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 97201  
QY 7749 TCGTGGCTCAAGGATCTCTCTGCTCAAGCTCCGCTATAGTGGACATCAAGCGCAT 7808  
Db 97200 TCGTGGCTCAAGGATCTCTCTGCTCAAGCTCCGCTATAGTGGACATCAAGCGCAT 97141  
QY 7809 AACACAGCTCGGCTCAATTTTGTATTTTGTGGTTACAAATGATATATATAT 7868  
Db 97140 AACACAGCTCGGCTCAATTTTGTATTTTGTGGTTACAAATGATATATATAT 97121  
QY 7869 TTAATTTGTATTTTGTATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG 7928  
Db 97120 ----TTTGTATTTTGTATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG 97065  
QY 7929 CTCTGACCTCAAGTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7988  
Db 97064 CTCTGACCTC--GTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 97007  
QY 7989 AGCAACAGGCTGCGGCTCAATAGTTTTATATATCTTTGCTTCTTCAACTTT 8048  
Db 97006 AGCAACAGGCTGCGGCTC--CGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 96949  
QY 8049 GCGTTCAGTGAAGCAGTATGTTCTCTCACTTCTCACTTCTCACTTCTCACTTCTCACTTCT 8108  
Db 96948 GCGTTCAGTGAAGCAGTATGTTCTCTCACTTCTCACTTCTCACTTCTCACTTCTCACTTCT 96889  
QY 8109 TCAGTTCAGTGAATCTCTCTCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8168  
Db 96888 TCAGTTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 96846  
QY 8169 GAATCTGAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8228  
Db 96845 GTGAGCAGTGAAGCAGTATGTTCTCTCACTTCTCACTTCTCACTTCTCACTTCTCACTTCT 96786  
QY 8229 GCGTATGTTAGTCACTTAATAGGTTGGAACCTCATCTTCACTTCACTTCACTTCACTTCA 8288  
Db 96785 TCTCGAATCTGATGATCAAT-----TATATCTCTCACTTCACTTCACTTCACTTCACT 96740  
QY 8289 AGCATAGCTCTGCTGCTCAAAATAGGCGCTGAGGCTGAGTGTGCTGATGCTTAT 8348  
Db 96739 GGAAGCTGAGTTAAGGCTGATATAGAAATTTTATAGGCAAGGCTGATGCTGCTGCTGCT 96680  
QY 8349 ATCCCAACATGTTGGAAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8408  
Db 96679 ATCCCAACATGTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 96622  
QY 8409 GCGTGGCAACATGTTGGAATCTCATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 8468  
Db 96621 GCGTGGCAACATGTTGGAATCTCATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 96563  
QY 8469 GTGGCATGACAGTATGCTCCAGCTGTTGAGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 8528  
Db 96562 GTGGCATGACAGTATGCTCCAGCTGTTGAGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 96503  
QY 8529 TGGGAGGAGAGTTCAGTGAACGTCGTGCTCTGCAATCCAGCTGCTGCTGCTGCTGCTGCT 8588  
Db 96502 CCGGAGTGAAGTTCAGTGAACGTCGTGCTCTGCAATCCAGCTGCTGCTGCTGCTGCTGCTGCT 96443  
QY 8589 ATTGAGACCTGCTGCAAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8648  
Db 96442 AGGAGACCTGCTGCAAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 96383  
QY 8649 GGTATTTGAGTGGAAACCTCCAGGAGGAGTGAATGTCAGTCAAGCAGCTGTATATCTC 8708  
Db 96382 CTTTATTAACCCCAACTATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 96323  
QY 8709 AGTACAGGCTTAATAGAGAACTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8768  
Db 96322 AGATGACTAATCTTATGATGCTGCTTCAAGATCACTGCTTATTTTCAAAAGATTT 96263  
QY 8769 AATATTTGTTGAGCATCTGATTAAGAGTGGGCAATGCTTAGGCACTGAGATACAGTACGT 8828

Db	96262	CAAAATATGATTATTTAAATAATTCGTGAATTTG-----GACAAATTTGTAACATGATT	96208
Qy	8829	CAACATGGCAGACAAAGATGCGCTGCGCAAGGCTCTGCTAAAGTGAAGAGACATTA	8888
Db	96207	CACATATCTTCAATCCATGCTCCT-----TGATAGTGAAGAAAAAAA	96168
Qy	8889	GAAGAAGAAAGAAAGAAAGAAATTAATTTTAAGGTAATTAATTAAGGTTGTAAAGAAATTA	8948
Db	96167	AAATATATAGTTGTAAATTCGTGTGCTTATATTTCTAGTGAATGTGTTCAATGCTAAAT	96108
Qy	8949	GACAGATTAAGTGAGATTAAGGTGAGAGAAATGAGGCTGTCTTCTGAAGAAATGATTTT	9008
Db	96107	AAATCTTTATATATATGCTCTTTGTAAACATCCTGTCCCTTTATCTTAAGAAAAATTA	96048
Qy	9009	GAGCTGAGACTTCAGTATAGAAAGAAATTAACCAACGATGTGCTGAGAGAAACATT	9068
Db	96047	TTCTTGGCAATTAATTAATTAATGTATTAATCTGCAATTAACCTTAACTTGTATCATATAT	95988
Qy	9069	TTAGGAGAGGTGAGCAGACATTACTTCAAGAAATCAAGAAAGACCTGAGGCTGGA	9128
Db	95987	TACTCTGCATTAACATTAATCTTGATACATTAATTAATCTTATATAGCCAGTTTAAAGAA	95928
Qy	9129	ACACAGAAAGAGCAGAGTGGTGACTTGAAGGGCAGGAGCGCAGTGGCCAGTTACC	9188
Db	95927	AGTGGGATCTTAATTTGTAAATAATGTATCAATGAAA-----ACTTTCAAGAACCA	95876
Qy	9189	TAGACTGTGTAAAGGTTTCAACCATTAAGGAGATCAAGAAATCTTGAAGCAGGCT	9248
Db	95875	GGGCGAGGCAAGATGGCTCACTCCTGTAACTTGAACCTTGTATAGCTGAGGCGAGAG	95816
Qy	9249	GTCATATATTTCTAACTCATTTTAAATTAAGATCACTGTGACTTTTTCAGAAACATAGT	9308
Db	95815	GTGATATCACTTGAAGCCAGAGATTAGAGACCAAGCTGGGCAAGGTAGAAGACCCCATCT	95766
Qy	9309	TATTAAGATACAGCATGTAAAGCAAGAAATCCAGCTAGCAATCCGTGACAGTTGTCCAAAT	9368
Db	95755	CTTTAAAGAAAAATTAATTTTATTTTAAATC-----TAAAGTTTTTTTAAAGA	95703
Qy	9369	TAAAGGTGAAGACCGCTTGAACCTAGATGATAGACAGACAGGCTGTGAAGAAATCACATG	9428
Db	95702	AAACTTTCAAGAACTTAATTTCTGTAGATTAATTAATTTCTTATATATAGATTGCAMAAT	95643
Qy	9429	ATATATTTTGAAGTGAAGCTGAACAGCATTAATCTAATAGCTAAGATAGAGCCGGG--TGTG	9486
Db	95642	GTGCACTTGAAGGAAAAATTAATTAATTTGTGTATGATTTTGAATCTGGGGCCAGGG	95583
Qy	9487	GTGGCTTAAGCCTGTAAATCTAGACACTTTGGAGGCCA--AGGCGAGTGAATCACTTGA	9543
Db	95582	GTGGCTATCCTGTAAATCCCAACACTGTGGAGGGCCACCGAGTGGGTGAATTCCTGA	95523
Qy	9544	GGTCAAGAGTTTCAAGACCAAGCTTGAACCAACATGTAAATCTCTGTCTAATTAATATAC	9603
Db	95522	GGTCAAGAGTTTCAAGACCAAGCTTGAACCAACATGTAAATCCCATTTCTAATTAATATAC	95463
Qy	9604	AAATATAGCTGGGAATGTGTGACCATGCTGTATCTACAGCTTATTTGGAGAGCTGAAGG	9663
Db	95462	AAATATAGCCGGGCGTGTGTGCTTAATGCTGTAAATCCAG-CTACTCAGAGAGCTGAAGC	95404
Qy	9664	AGAGAAATGCTTGAACCTTGGAGGTGAATGTTCAGTGAAGCCGAGATTCACATTTGA	9723
Db	95403	AGAGAAATGCTTGAACCAAGGAGGTGAAGTTGACAGTGAAGCAATTTGTCCACTGCA	95344
Qy	9724	CTCAGCCTGGGGAACAAAGATGAATCTCCGTCTTA	9760
Db	95343	CTCAGCTGTGGGCAAC-AGAGCAAGCTCCATCTCA	95308

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION
11 ACI08045/c	ACT08045	Homo sapiens BAC clone RP11-307L8 from 4, complete sequence.	ACT08045 AC021616	GI:27356779

KEYWORDS	HTS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Enkavaliota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 74904)
TITLE	Sullivan, J.E. and Waterston, R.
JOURNAL	Toward a complete human genome sequence
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
PUBMED	99063792
REFERENCE	9847074
AUTHORS	2 (bases 1 to 74904)
TITLE	Armstrong, J., Haglund, K., Courtney, L., Mangiapanello, L., Trani, L. and Biellick, L.
JOURNAL	The sequence of Homo sapiens BAC clone RP11-307L18
REFERENCE	Unpublished (2001)
AUTHORS	3 (bases 1 to 74904)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (24-JUN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	4 (bases 1 to 74904)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (23-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	5 (bases 1 to 74904)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (21-DEC-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	6 (bases 1 to 74904)
TITLE	Waterston, R.
JOURNAL	Direct Submission
REFERENCE	Submitted (27-JUN-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Dec 21, 2002 this sequence version replaced g1:1867310.
COMMENT	

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**  
The RPCT-II human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frensen, E., Tatenno, M., Catanesi, J. J. and de Jong, P. J. (1998) An improved

Approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is not the entire insert of the clone. This clone is overlapped by AC093940 and AC114735.

The sequence of AC021616 has been incorporated into AC108045.

#### FEATURES

##### source

1. 74904  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone\_lib="RP11-307L18"  
/clone\_1ib="RPCT-11"  
1. 272  
/rpt\_family="Alu"  
351. 650  
/rpt\_family="Alu"  
914. 1229  
/rpt\_family="Alu"  
1328. 1349  
/rpt\_family="AT\_rich"  
1492. 1786  
/rpt\_family="Alu"  
1789. 1811  
/rpt\_family="AT\_rich"  
1914. 2060  
/rpt\_family="L1"  
2061. 2376  
/rpt\_family="Alu"  
2377. 2432  
/rpt\_family="L1"  
2509. 2818  
/rpt\_family="Alu"  
2871. 3167  
/rpt\_family="Alu"  
4034. 4305  
/rpt\_family="Alu"  
4465. 4775  
/rpt\_family="Alu"  
4951. 5233  
/rpt\_family="Alu"  
5272. 5671  
/rpt\_family="L1"  
5672. 5951  
/rpt\_family="Alu"  
5952. 6133  
/rpt\_family="L1"  
6146. 6427  
/rpt\_family="Alu"  
6464. 6580  
/rpt\_family="L1"  
7263. 7541  
/rpt\_family="Alu"  
7774. 7903  
/rpt\_family="L2"  
8016. 8323  
/rpt\_family="Alu"  
8469. 8506  
/rpt\_family="AT\_rich"  
8507. 8793  
/rpt\_family="Alu"  
8910. 9210  
/rpt\_family="Alu"  
10007. 10307  
/rpt\_family="Alu"  
10308. 10327

```
repeat_region      /rpt_family="(T)n"
10696. 10964
/rpt_family="Alu"
repeat_region      11312. 11714
/rpt_family="L1"
repeat_region      11743. 13589
/rpt_family="SVA"
misc_feature        12319. 13447
/note="CPG island (%GC=71.6, o/e=0.70, #CPGs=81)"
repeat_region      13609. 13679
/rpt_family="L1"
repeat_region      13795. 14091
/rpt_family="Alu"
repeat_region      14106. 14263
/rpt_family="L1"
repeat_region      14285. 14596
/rpt_family="Alu"
repeat_region      14601. 14851
/rpt_family="L1"
repeat_region      14936. 15233
/rpt_family="Alu"
repeat_region      15270. 15559
/rpt_family="Alu"
repeat_region      15735. 15808
/rpt_family="Alu"
repeat_region      15805. 15854
/rpt_family="Alu"
repeat_region      15907. 16149
/rpt_family="Alu"
repeat_region      16669. 16771
/rpt_family="Alu"
repeat_region      16894. 16936
/rpt_family="AT_rich"
repeat_region      16970. 17123
/rpt_family="Alu"
repeat_region      17127. 17151
/rpt_family="(ATG)n"
repeat_region      17436. 17546
/rpt_family="L1"
repeat_region      17549. 17624
/rpt_family="MER1_type"
repeat_region      17748. 17903
/rpt_family="Alu"
repeat_region      17966. 18169
/rpt_family="MER1_type"
repeat_region      18174. 18276
/rpt_family="L1"
repeat_region      18277. 18564
/rpt_family="Alu"
repeat_region      18565. 18666

Query Match      4.7%; Score 471.6; DB 9; Length 74904;
Best Local Similarity 53.9%; Pred. No. 5,3e-114;
Matches 1420; Conservative 0; Mismatches 1044; Indels 172; Gaps 15;

QY 7169 AGAGAACCTTATATTAATAATACCCAGCGCGGTGAGTACCGCTGTATCT 7228
DB 2846 AGAACATCTCTCTATTAACAAATATAGCTGGGCTGGCTCAACACTGTATCCC 2787
QY 7229 AGCACTTGGAGGCGGAGGAGGAGTGGATGGCTTGAGCGGCTTGAGCGTGTGAG 7288
DB 2786 AGCACTTGGAGGCGGAGGAGGAGTGGATGGCTTGAGCGGCTTGAGCGTGTGAG 2737
QY 7289 ACCAGCTGGGAGCAATGCAAACTCTCTCAAAATAATTAATAATTAATGCTGGGT 7348
DB 2736 ACCAGCTGGGAGCAATGCTGAACCCCATCTCTAATAATAATTAATAATTAATGCTGGGT 2677
QY 7349 GTGTAGTGCCTTCTATATGTCCTATCTTCACTTCAAGAGCTGAG----- 7391
DB 2676 GTGTAGTGCCTTCTATATGTCCTATCTTCACTTCAAGAGCTGAG----- 7391
QY 7392 -CCCGAAGGTGAGGCTTCACTGAGCGGTGATGCTGCTACTGCACTCCAGCTGGGTGA 7450
DB 7450 -CCCGAAGGTGAGGCTTCACTGAGCGGTGATGCTGCTACTGCACTCCAGCTGGGTGA 7450
```

Dh 2616 ACCTGGAGGTGAGGTTCAGTGAAGCTTAGATTGCACTGCACTCCAGCTGGGGCGA 2557  
Qy 7451 CAGAGTGAGACATGTCTCAAAAAACAAAAACAAAAACAAAAACAAAAACAA 7510  
Db 2556 CAGAGTGAGACATGTCTCAAAAAACAAAAACAAAAACAAAAACAAAAACAA 2497  
Qy 7511 AAAAAAACCCTAT 7570  
Db 2496 ACAT 2452  
Qy 7571 AAAAAACCAACATGACCCAGCTACAGATGGGCAATCCCTCCCTCAGCTGTAAT 7630  
Db 2451 AGACAGAGCTCTTATATATATATATATATATATATATATATATATATATAT 2392  
Qy 7631 TTTTCTTCTCTGATCAGATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7690  
Db 2391 TTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2332  
Qy 7691 ACTCTGTACCCAGGCTGAGTGAATGGCGCAATCTTGGTTCATGCAACTCTGCTC 7750  
Db 2331 ACTCTGTACCCAGGCTGAGTGAATGGCGCAATCTTGGTTCATGCAACTCTGCTC 2272  
Qy 7751 CTGGGTTCAAGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7810  
Db 2271 CGGGGTTCAAGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2212  
Qy 7811 CACCATGCTGCTATATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 7870  
Db 2211 CACCATGCTGCTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 2194  
Qy 7871 AATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 7930  
Db 2193 -TTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 2136  
Qy 7931 CTGAGCTCAGGTGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7990  
Db 2135 CTGAGCTCAGGTGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2076  
Qy 7991 CAACCAAGCTGAGGCTCCTCATAGGTTTATTTATTTTATTTTATTTTATTTT 8050  
Db 2075 CCAAGGCTGAGGCTCCTCATAGGTTTATTTATTTTATTTTATTTTATTTTAT 2016  
Qy 8051 CTGCAAGTGAACCATATTTCTCTCACTTCTCTCACTTCTCTCACTTCTCTCACT 8110  
Db 2015 TAATTTTATTAAGAAATTCATTTTCAGAAATTTATAGAAATCAAGAAAGTACA 1956  
Qy 8111 AGTTTCAACATGATACCTCCCTCAGTGTCTCTTTTCTTCTGTAAGATTTTCAAGAGGA 8170  
Db 1955 AAGATGTCAT 1896  
Qy 8171 ATCTGAATGGCCAGTCCATATTTTTCAGACCAACCACTTAAAGTGTGATTCAGC 8230  
Db 1895 AGCTAA-----TAAATCACAGAAATCTATATATATATATATATATATATAT 1853  
Qy 8231 CTATGATTTGGCTACATTTATTTGGTGTGGAATCATCATTTACTTCAATGCAAGAG 8290  
Db 1852 CAAGGCTGAT 1793  
Qy 8291 CATAGCTGTGTTCTCAAAATAGGGCCCTGGGCTGAGTGTGTGTGTGTGTGTGT 8350  
Db 1792 TATATCA-----GGCAGGCTCAGTGGCTCAGCTGTATAT 1758  
Qy 8351 CCAACA-CTGTGGAGGCGAGGGGGGAGATCATTTGATCAGAGTTCTAGAACAG 8409  
Db 1757 CCAACAGCTTTTGGAGGCGAGGCGAGGAGATCTTTGAGGCGAGAGTTCAAGACAG 1698  
Qy 8410 CCGTGGCAACATGCTGAATCTCATCTCTATATATATATATATATATATATATAT 8469  
Db 1697 CCGTGGCAACATGCTGAATCTCATCTCTATATATATATATATATATATATATAT 1639  
Qy 8470 TGGCATGACACATGATGCTTCAAGGCTTCAAGAGCTGAGGTGGAGGATTTGCTGAGTGT 8529  
Db 1638 TGGCATGACACATGATGCTTCAAGGCTTCAAGAGCTGAGGTGGAGGATTTGATTC 1579

Qy 8530 GGGAGCAGAGATTTGAGTGAACCTGTACTGTCTGCAATCCAGCTGGGTGACAG 8589  
Db 1578 AGAGAGTGAAGTTGAGTGAACCTGTACTGTCTGCAATCCAGCTGGGTGACAG 1519  
Qy 8590 TTGAGACCTGTCTCAAAAAACAAATTAATTAATTAATTAATTAATTAATTAAT 8649  
Db 1518 GTGAGACTGTGATTAATAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1480  
Qy 8650 GTAAATTTGAGTGAACCTCCAGGAGGTGATATGTCAGTCAACCGCTATATCTCA 8709  
Db 1479 ACATATTTCAAGTAAAGAAAGCTTAAGAAATTTTCTATTTTCTCATGTTTATACA 1420  
Qy 8710 GTACAGGCTTAATTAAGAACTTGTGTGACAGCAAGAACTATAGGATTTTATCAACA 8769  
Db 1419 TGAAGAAATCATGCTGATGATCATATAGAAACACTTAAGACTGAATGCAAGCAGTT 1360  
Qy 8770 ATATTTGTGAGATCTGATTAAGAGTGGCATTTGCTAGGCACTGAGATACATATGTC 8829  
Db 1359 GAGAAATTTGTTTAATTAATAATTAATAATTAATAATTAATAATTAATAATTAAT 1300  
Qy 8830 AACATGGCAGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8889  
Db 1299 TAAAGACACAT 1240  
Qy 8890 AAGAGAAAGAAAGAAAGAAATTAATTTTATAGTAAATTTTATAGGTTTAAAGAAATTAAG 8949  
Db 1239 GTTACTTTAGGCTGGGCACTAGTGTCTCAAGCTATATATCCACACTTTGGAGGCCAA 1180  
Qy 8950 AAGAGATGAGGATTAAGAGTGAAGAAATGAGGCTGCTTCTGAAAGAAATTTTGTG 9009  
Db 1179 CGCAGCAGATTAATCTGAGGTGAGGATTTGAGAGAGCTGCTGCAACATGCAAAACC 1120  
Qy 9010 AGCTGAGCTTCA-----GTATGAGAAAGAAATTAACACACATGCTGAGAGAAAGC 9065  
Db 1119 CGTCTCTATTAATAAAACAAATAATTAAGCAGCATAGCAGGGGCTGTTATCTTAG 1060  
Qy 9066 ATTTTGAAGAGGTGAGCAGCATATCTTCAAGAAATCAAGAAAGAAAGCTGCTGAGGCT 9125  
Db 1059 CTACTCTGAGAGGTGAGGCAAGAAATCGTTGAAACCAAGAGGTGAGGTTGACATGAG 1000  
Qy 9126 GGAACACAGAAAGAGCAGGTGGTGAATTTGAAGGGCAGGAGCGC-----AGTGC 9179  
Db 999 CCGAGATGCGACACACACTCCAGCTGAGGAAAGAGAGGATCTCATCTCAATAATAA 940  
Qy 9180 CAGTTACCTAGCCTGTAAAGGTTTTCACATTAAGGAGATCATCAAGAAAGCTTG 9239  
Db 939 TAAATTAAGTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 880  
Qy 9240 AGCAGGCTGTGATATATTTCTAATCTATTTTATTAAGATCATCTTGTGAG 9299  
Db 879 GAAATTAAGTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 820  
Qy 9300 AACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9359  
Db 819 AAAAGCTAACACACAGAGCTATTTTCACTGAGCTTAATTTCTTATGACATATATCTTT 760  
Qy 9360 TGTCCAAATTAAGAGTATGAGCCCTGAGCTAGATGATGATGAGAGAGAGGTGAGGA 9419  
Db 759 TGGCAAAATTTTATTAAGAAATTAAGCACTGATATATCAATTTTATCAATCT-----T 707  
Qy 9420 ATCAACATATATATTTTGGAGTGAAGCTGACAGATTTAATTAATTAATTAATTAATTAAT 9479  
Db 706 ATATATTTTAAGGAAATATGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 647  
Qy 9480 GGGTGTGTGCTTACGCTGTATATCTAGCACTTTTGGAGGCGCAAGCGAGTGAATCAC 9539  
Db 646 AGCGGCGGTGCTCATGCTGTATCCAGCACTTTTGGAGGCTGAGGAGGTGATCAC 587  
Qy 9540 CTGAGGCTCAGAGTGTGAGACCAAGCTTGAACCAATGATGAACCTGCTCATCAATAAA 9599  
Db 586 CTGAGGCTCAGAGTGTGAGACCAAGCTTGAACCAATGATGAACCTGCTCATCAATAAA 527



[illegible]

Accession	Sequence	Position
Dh	95467 AAAATTAGCCGGGCGTGTGGTCCCTTAAGCCTGTAATCCAG - CTACTCAGAGAGCTGAGGC	95409
Qy	9664 AGAGGATCGCTTGAACCTGGGAGGTAATGTTGCAGTGAAGCCGATTTGCACCAATTGA	9723
Dh	95408 AGAGGATCGTTTGAACCAAGGAGGTGGAAGTTGCAGTGAAGCCAGAAATTGTGCCACTGCA	95349
Qy	9724 CTCGAGCCTGGGAGACAAAGATGTAATACTCCGCTCTTA	9760
Dh	95348 CTCGAGCTCTGGGCAAC-AGAGCCAGACTCCATCTTCA	95313

RESULT	13
LOCUS	AC021105
DEFINITION	187360 bp DNA linear
ACCESSION	AC021105
VERSION	AC021105.13
KEYWORDS	GI:14589664
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 187360)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)

REFERENCE 2 (bases 1 to 187360)  
AUTHORS Sun, H., Abbott, A., and Spaulding, L.  
TITLE The sequence of Homo sapiens BAC clone RP11-163017  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 187360)

**TITLE** Direct Submission  
**JOURNAL** Submitted (14-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
**REFERENCE** 4 (bases 1 to 187360)  
**AUTHORS** Waterston,R.H.

**JOURNAL** Submitted (03-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

**REFERENCE** 5 (bases 1 to 187360)

TITLE	Direct Submission
JOURNAL	Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Jul 3, 2001 this sequence version replaced gi:13431110.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watsen.wustl.edu](mailto:sapiens@watsen.wustl.edu)  
----- Summary Statistics -----  
Center project name: H\_NH0163017

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

## VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-163017; actual end is at base position 187360 of RP11-163017.

The region between bases 32935-33083 is single stranded and the sequence fidelity can not be guaranteed. RP11-163017 contains an unresolved dinucleotide repeat between bases 47950-48140 where the sequence fidelity can not be guaranteed. Digest information suggests that approximately 88 bases are missing from the current assembly.

## FEATURES

## source

1.187360 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-163017"  
 /clone\_lib="RPCT-11"  
 1.85  
 /note="similar to EST AA397735 (NID:g2050626) zc86a05.r1"  
 misc\_feature  
 1.85  
 /note="similar to EST A1178109 (NID:g3728747)"  
 misc\_feature  
 1.85  
 /note="similar to EST AUI20774 (NID:g10936009)"  
 misc\_feature  
 1.85  
 /note="similar to EST AV610323 (NID:g9745993)"  
 misc\_feature  
 1.85  
 /note="similar to EST BB074520 (NID:g8422511)"  
 misc\_feature  
 1.85  
 /note="similar to EST BB614739 (NID:g9896336)"  
 misc\_feature  
 1.85  
 /note="similar to EST BE786771 (NID:g10207969)"  
 misc\_feature  
 209.800  
 /note="similar to EST BG432741 (NID:g13339247)"  
 misc\_feature  
 227.725  
 /note="similar to EST BP911522 (NID:g12302980)"  
 misc\_feature  
 649.800  
 /note="similar to EST AA397735 (NID:g2050626) zc86a05.r1"  
 misc\_feature  
 649.800  
 /note="similar to EST A1178109 (NID:g3728747)"  
 misc\_feature  
 649.800  
 /note="similar to EST AUI20774 (NID:g10936009)"  
 misc\_feature  
 649.800  
 /note="similar to EST AV610323 (NID:g9745993)"  
 misc\_feature  
 649.800  
 /note="similar to EST BE614739 (NID:g9896336)"  
 misc\_feature  
 649.777  
 /note="similar to EST BE786771 (NID:g10207969)"  
 misc\_feature  
 649.777  
 /note="similar to EST BE074520 (NID:g8422511)"  
 misc\_feature  
 651.800  
 /note="similar to EST BG703239 (NID:g13975373)"  
 misc\_feature  
 655.800  
 /note="similar to EST T32984 (NID:g615082)"  
 repeat\_region  
 1023.1310  
 /rpt\_family="Alu"  
 repeat\_region  
 1309.1334

repeat_region	/rpt_family="TTAA)n"
1678.1878	/rpt_family="MIR"
1938.1965	repeat_region
2379.2416	/rpt_family="AT_rich"
2399.2711	/rpt_family="AT_rich"
3034.3091	repeat_region
3295.3779	/rpt_family="MIR"
3784.3834	/rpt_family="LI"
3913.4049	repeat_region
4261.4340	/rpt_family="MIR"
4261.4340	misc_feature
4261.4340	/note="similar to EST A1178109 (NID:g3728747)"
4261.4340	/note="similar to EST AUI20774 (NID:g10936009)"
4261.4340	/note="similar to EST AV610323 (NID:g9745993)"
4261.4340	/note="similar to EST BE614739 (NID:g9896336)"
4261.4340	/note="similar to EST BG432741 (NID:g13339247)"
4261.4340	/note="similar to EST BG703239 (NID:g13975373)"
4261.4340	/note="similar to EST T32984 (NID:g615082)"
4261.4318	misc_feature
5010.5100	/note="similar to EST AA397735 (NID:g2050626) zc86a05.r1"
5163.5469	repeat_region
5163.5469	/rpt_family="L2"
5165.5191	repeat_region
5503.5802	/rpt_family="Alu"
6252.6283	repeat_region
6420.6447	/rpt_family="AT_rich"
6840.6916	repeat_region
6917.7061	/rpt_family="MIR"
7062.7376	repeat_region
7347.7388	/rpt_family="Alu"
7377.7412	repeat_region
7413.7448	/rpt_family="Alu"
7434.7465	repeat_region
7802.8097	/rpt_family="AT_rich"
8200.8459	repeat_region
8604.8915	/rpt_family="Alu"
8886.8917	repeat_region
	/rpt_family="(A)n"

Query Match 4.4%; Score 444.8; DB 9; Length 187360;  
 Best Local Similarity 53.5%; Pred. No. 7.7e-107;  
 Matches 1431; Conservative 0; Mismatches 1122; Indels 123; Gaps 19;  
 QY 7158 CCACGACATGAGAGACCTTATATTAAATACCCAGCGCGCGTGTGAGTCAAG 7217



Db 54129 CCATGCCCTGCAATGATCCCTGATTGAAACAAACAGGAGGCCAGAGTAGAGGCTGACG 54188  
Qy 7218 CCGTATACCTACACCTTTGGGAGGCTGAGCAGGTGATTTGCTTGAAGCGCTTGAGCCT 7277  
Db 54189 CCGTATACCTACACCTTTGGGAGGCTGAGCAGGTGATTTGCTTGAAGCGCTTGAGCCT 54241  
Qy 7278 AGGAGTTTGAAGCAGCCTGGGCAACATGACAAACCTCATCTTCAAAATAA- TAA 7236  
Db 54242 AGGACATGTATGACAGCTGGCTTAACTGTTAAACCTGCTTAACTAAATAAATAA 54301  
Qy 7337 AATTAGTCGGGTGTGATGTCGTTCTTATATGTCCTTACTTACAGAGC----- 7387  
Db 54302 AATTAGTCGGGTGTGATGTCGTTCTTATATGTCCTTACTTACAGAGCAGG 54361  
Qy 7388 -----TGACCCCGGAAGGTGAGGCTTCACTGAGCCGTATGTCCTTACTGACCTC 7438  
Db 54362 AGAATTGCTTGAACCTCGAGGAGAGAGTTGCAGTGAATGAGATGTGTCACATCACTC 54421  
Qy 7439 CAGCCTGGGTGACAGAGTGAACCATGTCCTCAAAAAACAAAAACAAAAACAA 7498  
Db 54422 CAGCCTGGGTGACAGAGTGAACCATGTCCTCAAAAAACAAAAACAAAAACAA 54481  
Qy 7499 AACAAACAAACAAACAAACAAACCAATATATATATATATATATATATATATAT 7558  
Db 54482 AA-GCATGTGTATCATTAAGACAGTCATCTATATCTGTTCTGTATGTTATGAG-CTATC 54539  
Qy 7559 CACTATTTTGTATTAATCAACCAATGACCCAGTCACAGATGAGGAGGAGCCCTCCCTC 7618  
Db 54540 CTGGCTCTGGAATATGCTCTGCTATTAACATGACCTGTCGCAACCAATCAACCTTT 54599  
Qy 7619 TCATGTTTAAATTTTCTTCTGACTCAGATTTGTTGTTGTTGTTGTTGTTGTT- 7677  
Db 5460 TTTAAAAAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTT 54659  
Qy 7678 -----GAGATGAGTCTCACTCTGTCACCCAGCTGAGTGCATGCGCAATCTTGG 7730  
Db 54660 CCCCCAGAGACAGATCTTACCCAGTCACCCAGGCGAGAGTGCAGTGCATCTATC 54719  
Qy 7731 TTCACTGCAACCTCTGCTCTGAGTTCAAGCGATCCCTGCTGCTGCTGCTGCTGCT 7790  
Db 54720 CTCACTGCAACCTCTGCTCTGAGTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCT 54779  
Qy 7791 CTGAGCATAGGCGCATATCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7850  
Db 54780 CTGAGCATAGGCGCATATCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 54832  
Qy 7851 AATGACTATTTATTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTAT 7910  
Db 54833 -----TTTGTAGAGAAAGGCTTACCATGTTGT 54862  
Qy 7911 GGGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7970  
Db 54863 GGGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 54922  
Qy 7971 GCTAGATTAAGGATGAGCAACAGCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 8030  
Db 54923 GTTGGATTTACAGGTGAGCCTGAGCCTGAGCCTGAGCCTGAGCCTGAGCCTGAGCCT 54978  
Qy 8031 TTGCTTCTTCAAACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8090  
Db 54979 TTGCTTCTTCAAACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 55038  
Qy 8091 CTTTCAATGATCTTCAAGTCTGATTTCAATGATCTTCAATGATCTTCAATGATCTTCA 8150  
Db 55039 AACAGACAGATTAATAAGTATTTCTTGAACAAAGAACTTCAAAATATATAGAACAA 55098  
Qy 8151 GTAGATTTTCAAGAGAGGATCTGAATGAGCCAGTGCATATTTTCAAGCAACCAT 8210  
Db 55099 AATGTGCAAGAGGAGAGTGAATAAATCAACCATATTTTGGGATTTTAAACCC 55158  
Qy 8211 TAAAGTGTGATTTGCAAGCTATGATTTGCTATCAATTAATGAGTTGGAATCATCAT 8270

Db 55159 TAAATTTTATTAAGATATCAAAAGATCTGAATTAATCTATCAACCAAGCTTTTGTACT 55218  
Qy 8271 TACTTATTTGCAAAAGACAGATAGCTGTGTTCTCAAAATAGAGCCCTGCGCCAGTG 8330  
Db 55219 GACATTAATGATATCTACCTTACG- - -ACTATTTAAAAAAATTTCTTGGCTGGCA 55275  
Qy 8331 TGGTGGCTGATGCTTAATTAATCCCAACATGTTGGAGGCGAGAGGCGAGATCATTTGAG 8390  
Db 55276 AGGTGCTATGCTGTATATCCAGCATCTCGGAGGCGCAAGAGGCGGATGCTTAAG 55335  
Qy 8391 TCCAGAGATTTAGACACCACTGGGCAATGATGAAATCTCATCTCTAATAAATAC- 8449  
Db 55336 GTCAAGAGTTTCAAGACCACTGGCCCAATGATGAAACCTGTCTTCTAATAAATAA 55395  
Qy 8450 AAAAAATTAGCAGAGTGTGGGCAATGACACAGTATCCAGCTGTTTACAGAGCTGAGG 8509  
Db 55396 AAAAAATTAGCAGAGTGTGGGCAATGACACAGTATCCAGCTATCTTGGAGAGCTGAGG 55455  
Qy 8510 TGGAGGATTTGCTCGAGTGTGGAGGAGAGATTTGAGTGAACCGTACTGTGCTCTGC 8569  
Db 55456 CAGGGGATTTGCTTGAACCGGGAGGAGGAGTGTGAGTGAAGCCAGATTGCGCCACTGC 55515  
Qy 8570 AATCCAGCTGGGTGACAGATTTGAGACCTGTCTCAAAAAACAAATTAATTAATTAAT 8629  
Db 55516 ACTCCAGCATGGGTGACAGGCGAAGCTGTCTCAAAAAACAAATTTCTTTTAAAT 55575  
Qy 8630 AATATGTTCTTCCAGCAGGATTAATTTCACTGAGGAAACCTCCAGGAGGAGTATGT 8689  
Db 55576 GCACAGAGAGATTTCAACAG-AGAGACCGTATATCTGATCAAAAAATTAAGTCTAAT 55634  
Qy 8690 CAGTCAACGCTGTATATCTAGATACAGGCTAATTAAGAACTTGTGTAGACAGCAAAAC 8749  
Db 55635 TCATATGATGAGATTTTAAAGATATTAATGTTCTCTGCAATATCATATATTAATTA 55694  
Qy 8750 ACTAGTATTTACTCAACAAATTTTGTGAGATCTGATTAAGAACTGGGCAATGCTTA 8809  
Db 55695 AATCAATTAATTAAGATATGATTAAGAAATCTCAAGTGTGGGAAT-TAAATGGCAAC 55753  
Qy 8810 GGCACTGAGTACAGATGCAACATGAGCAGACAAATGCTGCTGCTGCTGCTGCTGCT 8869  
Db 55754 TTCTTTATTTATCTATGCTGAGTAAATCAACAGCAATTTAGAAATTTGAAAGTGA 55813  
Qy 8870 AAGTGAAGAGCAATTAAGAAAGAAAGAAAGAAAGAAATTTAGTATATTA 8929  
Db 55814 ATGATATTAATAAGATTAATTAAGAAATTAAGAAATTAAGAAATTTAGGAAAGT 55873  
Qy 8930 AGGTTTAAAGAAATTAAGACAGATAGTGGATAGAGTGAAGAGATGAGGCTGTC 8989  
Db 55874 TTCTATTTAAAGCTTACTGTAAAGAAAGAAAGTAAATCAATCAATCAATTA 55933  
Qy 8990 TTCTGAAGAAATTTTGAAGTGAAGTTCAGTATGAGAAAGAAATTAACACAGAT 9049  
Db 55934 CAGAGGATGACAGATTTTGTGTAAAGTCCAGTATGCAAAATTTGAGAGCTTTGTA 55993  
Qy 9050 GTGCTGAGAGAAAGCAATTTTGAAGGAGTGAAGCAGACATTTCAAGGAATCAAGAG 9109  
Db 55994 GTATGTAAGTGTGATATTT- - -CCCCCTGCCCCAACATTTAAATATTAATATCAG 56051  
Qy 9110 GAAAGCTGTGAGGCTGAGAACAGAGAAAGAGAGAGTGGTGAATTTGAAGGCGAGGA 9169  
Db 56052 GCCAGGCTGATGAGTGCATATGCTGTATCCAGCACTTTGGAGTGTGAGGCGAGTGA 56111  
Qy 9170 CGGCAATGGCAGGTTACCTAGACCTGTA-----AGGTTTTCACACATTAAGAGGA 9222  
Db 56112 CAGTTGATGACAGATTTTGAAGCAGCCTGACCAACATGATGCAACCCATCTGTGTA 56171  
Qy 9223 GTATCAAGAAAGCTTTGAGCAGGCTGTGATATTAATCTAATCTATTT-----T 9271  
Db 56172 AAAAAATTAAGAGCGGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 56231  
Qy 9272 TATTAAGATCACTGATCTTTTGTGAGAACATTAATTAATTAAGTGAAGATGTAAGC 9331  
Db 56232 GAGCAGAGAGTGTGCTTGAACCTGGAGAGCAGAGGTTGAAGTGAAGCAGATCAACCA 56291



QY	9332	AAGCAATCCAGC-----TACCAATCCGTGACGTTGTCGAAATTAGAGGTG	9376
Db	56292	CTGCATCTCCAGCCCTGGGTGACAGAGTAGGCACTGTCTAAAAACAAACAAAGAAACAA	56353
QY	9377	ATGACCCGCTTGAGACTAGATGATAGCAGAGGTGTGAGAAATCAACATGATATATTT	9436
Db	56352	CAAAAAACCAATTTTATAGTAGGAAGCAATATATTAAGACATAGGCGATCATTTAGATTT	56411
QY	9437	TGAGAGTAGAGCTGACAGCATTTAACTTAATAGCTAATAGTACGCGGGGTGTGTGGCTTAGG	9496
Db	56412	TGGG--AACCCGTATCTGTGTTCATCTTAAGATTTAGCCGGGGCGGTGGCTCACCC	56465
QY	9497	CTGTAACTCTAGACATTTTGGAGGCCAGCGAGGTGATCACTGAGTCAAGAGTTGG	9556
Db	56470	GCTGTGATCCAGACATTTGGAGGGCGAGGACGGGGATCACTGAGGTCTGGAATTTGG	56523
QY	9557	AGACCAAGCTTGACCAATATGTGAAACCTGTCTCTAATAAAATAC--AAATATAGCTG	9614
Db	56530	AGACCAAGCTTGACCAATATGTGAAACCCGTCTCTAATAAAATACAAATAATTAGCCA	56588
QY	9615	GGAATGGTGGCACTAGCTGTAAATCTCAGGCTACTGGGAGGCTGAGGACAGAGATGGC	9674
Db	56590	GGGGTGTGTGGCGATGCTCTAATATCCAG--CTACTTGGAGGCTGAGGACGAGAAACCG	56648
QY	9675	TTGAACCTGGAGGTGAATTTTGACGTAGACCGAGATTTGACCACTTGACCTCAGCTGG	9734
Db	56649	TTGAACCAAGGTGGCGGAGGTGTGACAGTGAACCAAGATTTGACCACTTGACCTCAGCTGG	56708
QY	9735	GGAACAGAGTGAACCTCCGTCTCTAATAAATGAA	9770
Db	56709	GCAACAGAGTGAACCTCCATCTCAAAAAAAAAA	56744
RESULT 14			
HSBA12M9/c			
LOCUS	HSBA12M9	176287 bp	DNA linear PRI 04-MAR-2003
DEFINITION	Human DNA sequence from clone Rpl1-12M9 on chromosome 22, complete sequence.		
ACCESSION	AL080243		
VERSION	AL080243.21	GI:58706222	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1. (bases 1 to 176287)		
TITLE	Pelan,S.		
JOURNAL	Direct Submision		
COMMENT	Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 12, 1999 this sequence version replaced gi:5777540.		
	----- Genome Center		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web Site: http://www.sanger.ac.uk		
	Contact: humquerry@sanger.ac.uk		

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22> Rpi1-1249 is from the library RPCI-11.1 constructed by the group of Plietser de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6. (Genbank accession

FEATURES	SOURCE
Location/Qualifiers	1. 176287
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="RZPD:RPCIB753M0912"	
/db_xref="taxon:9606"	
/chromosome="22"	
/clone="RP11-12M9"	
/clone_1b="RPC1-11.1"	
BASE COUNT	45020 a 40451 c 40308 g 50508 t
ORIGIN	

Query Match	4.4%	Score 439.6;	DB 9;	Length 176287;
Best Local Similarity	61.8%	Pred. No. 1.9e-105;		

Matches 909; Conservative 0; Mismatches 484; Indels 79; Gaps 10;

Qy 7182 TATTAATAATACCGAGCCGGGCGTGTGAGTACAGCCGTATATCTAGACATTGGGAG 7241

Db 130093 TTTTAACCCCAACGCGAGGGGTGTGCTACAGCCGTATATCCCAACATTGGGAG 130034

7242 GCTGAGGCGAGGTGGATTGCTTGAAGCGGCTTGAGCCTHGGAGTTTGAACCCAGCCTGGGCA 7301

Db 130033 GCCGAGGTGGTGAT-----CATGAGTCAGGAGTCAAGACCAGCCTGCCA 1299

Qy	Db
7302	129984
ACATGGCAAAACCTCATCTCTACAAATAATATAAAAATTAGCGGGGTGTGATGCGTT	ACATATGTAACCCCATCTCTACTAAAAATACAAAAATTAGCGGGCTGTGTGTCATGTG
7361	12999

[illegible]

Db 129924 CCTATAGTCCCACTACTCAGGAGGCTGAGACGAGAGATCCGCTTGAACCCGGAGGCGG 1298

7404 AGGCTTCAGTGAGCCGTGATCGTCTACTGCACCTCCAGCCTGGGTGACAGAGTGAGACCA 7463

Dy	7664	TGTCTCAAAAAAAAACAAACAAACAAAACAAAAAACC	7523
Dd	129864	AAGCGTGAAGTGAGACTGCACCACTGACTCCAGCTGGATGACAGAGTGAGACTC	1298

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

QY 7524 TATATATATATA-----CCTAGCTGAGTGAGATGCACTATTTGTGAAA 7573

Db 129744 TTTT'TTTTGAGATGCAGT'TTCACTCTCGTGGCCAGGTGCAATCTCAGCTCATGGCAACC 1296

7574 ATCACCACCATGATCCCGACTACGACATGGGGAGTCCCTCCCTCTACCTGTA-- -AATT 7631  
 129664 TCCGCTCCCGGGTTCAAGAGATTCCCTTGCCTCAGCTCCCGGAGTACGCTGGGACTACAG 1296

7632 7691

Db 129624 GGCCTGACACACGCCCGCTTTTTTTTTTTTTTTTTTTCTGAGATGAGCTCG 1295

7692 CTTCTGTCACCCAGGCTGGAGTGCATCTTTGGTTCATCTGCACCTCTGGCTCC 7751

DB	129564	CTCTGTGACCTGAGTTCAGTGGACACATCTCCGCTACCTGCTAGCTCCGCTCC	1295
QY	7752	TGGGTTACGCACTCCTCTGGCTCAGCTCCCGTATAGCTGGGACATACAGGCGCATACC	7811

Db 129504 TGGGTTACGCCATTTCTCCTGCTCAGCCTCCTGAGTAGCTGGACTACAGGCGCCGCC 1294



Dh	2490	AGCAGTCCTCCCAATTAATGATTTTACACTCTTTCTTACAAAGCAGCTTTCTT-CTGTGTTTC	2548
Qy	3584	CTGTTTCACTAATAATGACCCCAAAATGTGCTCGATTAATTACAAACCAAGTCACACAGGCT	3643
Dh	2549	CTGATTTCACTAGTAGA-GGTACCACTCGGTACTTGGGATACCATGTTCCAGGTGTACA-----	2601
Qy	3644	CTCATGATCTGGGGCTTGTTATCTTCTCAGGTTTATCTCTCCCTGCGCACAATTCACTG	3703
Dh	2602	-----TGGGGCTGAGGAGCGCTATTTCTGTGCTTTTT-----CTTCCATGTTTCACTG	2646
Qy	3704	TGTGCGACGCACTTACGAATCTCACTATGAGGTTGAGAGCAGCTGTTCTCTCACTGTTGGGCTC	3763
Dh	2647	TGTACAGCTGCCATTAAGTTTATCCAAAG-TATCCGAATACTTTATATCATGTGTACAGTCC	2705
Qy	3764	TGCATGCTGCTCCCTCTGTGCTGTAAACACCTTTCTCTCACTTGTCAACCTGGAATAATTCCT	3823
Dh	2706	TCTGTGCTGCTGTGCTC-----CTTGGAAATAATCT	2733
Qy	3824	GCTGATTTTTTACGCTCTTGGGGCCCAATGCTTCTCTTTGGTGTAAACCTTCCACAATT	3883
Dh	2734	CACATCATTTTCAATCTTGTGGCAATATCTTCTCTCCACTGTAATACTTTCTGCGGGCTT	2793
Qy	3884	CTCTAGGCACTTATAGGCACTGTGCTATATCTTCAAGTCACTGCTTTACAATAACCTGTG	3943
Dh	2794	TTCCAGACATTAATTAAGACATTTCCCCCAATATTCATATGCTGTTGATATTT---CTTTT	2850
Qy	3944	GTAAGTTGATGGCTAGATTTGACAGAGTCCTTTGTGCTTTTGTACAGTGAATCTTCTGAA	4003
Dh	2851	TTATCTTGTTGGTGTGAACGGCTGGGGAATGTATGTTAGTTTTCAGGGAATCTTGCTCA--	2908
Qy	4004	GTGAAGAAGACAGTCTTGTATATCTCACTGCTCTCAACATGCTCGGCATATAGTAGTTA	4063
Dh	2909	-----TTCCCGAATGTTCAACGGCTCTGTGATGCTGAGGCTGTACTAGTCA	2955
Qy	4064	TTCACTGACGTGTTCTTGGAGATGATGATGATGATTAATAATTAATGAAGAAATGAATGAA	4123
Dh	2956	TTACGCGGTGTTTCTTGAAAGAAATCGTAATATGATGGCAAAATGAATATATATAATGAGG	3015
Qy	4124	GAATTAACGTATGGGTGATATGACAGATGAACAGTTGTGATATGTTGTCAACCTGATA	4183
Dh	3016	CAAT-----GTATTTAGAGGATTAATCAATGTGGATGTGCTCGCCAAAATCTAAGA	3064
Qy	4184	GTGTTGCATTAATATGTGCCACAGGAGTGTCTGG--GTACAGACCTAGAGGCATGTGTGT	4241
Dh	3065	GAGCTGTATTAATATGTGTCACAGGCGCATGTCTGTTATACATPACAGAGGTGTATGTGAT	3124
Qy	4242	TATATGTAATATGTAAGTGTGATTTTGACAAACCTGAGAGTGTGTATATGTGCAAAAGACAGA	4301
Dh	3125	TCTTAGTAGGATGGCTGTGGTTTG-----CTGAAGATATGTATGTATGAC-AAGGGGACACA	3177
Qy	4302	CATTGTTGTCCACAGATGGACTGAGAAATGTGTAGGGCCACAGAAAGGATATGTATTAAGA	4361
Dh	3178	TGCTGTACTCACCCTGACGCTGTGAATAATGATATGCTGTAGGAAGATATCTGTCTTAAGCA	3237
Qy	4362	CAGTAGATTAATAAATATGTGTGTAAATGCAAGATGGCAATATCTGGGGATGCAACGTCAAA	4421
Dh	3238	CAATATAGACAAATAGTCATGTGAATGCACTGCATCTGAGGTGAACGGTTCCAAAGACAG	3297
Qy	4422	AG-----AGAGTACTTTGAATGACAGGG	4445
Dh	3298	TGCTTTGTTGTTTTTTTTTTTGTGTTTTTTTGTGTTTTTTTTTTTTTTTGTGCTCAGGAA	3357
Qy	4446	GACAAAGCTGGGTATACCTCTGTGAAGAAAGAGAAAGGATPCCCAAAATGCTGCCAA	4505
Dh	3358	GACAAAGTCTGGGTGTACCTCTCTGTGAAGAAAGAAACGGGATATCCAAATTTGTTCCAA	3417
Qy	4506	GATGAATTTCTGGAATCCATCCCACTGAGATGCACTGTGATGTCAACCTGTGAGAA	4565
Dh	3418	GAAAGATTTCTCGAATCCCATACCCACACAGACGAGCTGTGGAATGCAACGTGGAGAG	3477
Qy	4566	TGTCAAGCTGAAGCGCTCTCTTTCTTCCAAAGTCAAGTCCAGGCTGTGAATCCAAAGACTG	4625
Dh	3478	TGCCAGAGTGAAGCGTCTTTTCTTCCAAAGTGTGCTTACACAGAAACCAACAC---	3534

[illegible]

Db	5486	AAAT-----AGAAATACCTTTTAACTTCTAATGACTAAGCAAAATATGATATGTTCTG	5538
Qy	6752	ACTATCAAGTTCTGTGAAAATATCATCACCATGAGAGCAAAATCTCCACAGCAGAGT	6811
Db	5539	ATTAAACAAGTCTGTGTATAAGTATGCACTACATCGAATGTAGCT-----CAACATGAA	5593
Qy	6812	TGCAACATATAATAGAAACATACAGCTAAGTGAACACACACCTGTAGTAAATACAA	6871
Db	5584	TGCACACCTTATATGAAACATACAGCTGCATTAATAACATGAA--TGCTGTACAAATGAG	5651
Qy	6872	CATTAACTAGAACATACGCCCATAGTAAAGAACATATGATATCA-----AGAGAACACA	6936
Db	5652	ATCTGAGCT--GATTATATGTCAATATTTAAACATGTAAAGTCAAGATTTATGAGTCTA	5709
Qy	6927	CAGCCATGTGTGGAGGCCATTGGAGAGACACACGACAAATGTGAATGCGAAMAAAGAGA	6986
Db	5710	TAGTAAGTGTGTAGATCACTGGGGAAACACACAGACAAAGTAAGTATGTGTGTGTGT	5769
Qy	6987	GAGAGTAGTGAAGAGATTGTGAAAACAGGGCCACAGGAAACACACAGAAATAGAGAGAG	7045
Db	5770	GTGTGTGTGTGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	5828

Search completed: December 15, 2003, 23:41:40  
Job time : 23561.6 secs

```

Db      5486 AAT-----AGAAATACCTTTTAACTTAACTAAGCAAAATATGATGTTGTG 5538
Qy      6752 ACTATCAAGTTCTGTGAAAAATACATCACCACTGAGAGCAAAATCTCCACAGCAGGAT 6811
Db      5539 ATTAACAAGTCTAGTGAAAAAGTATGCAACTACATGGAATGTAGCT-----CAGCATGAA 5593
Qy      6812 TGCACACTATATTAAGAACTACAGCTTAAGTGAACACACACCTGTAGTGAATAATCAA 6871
Db      5594 TGCACACCTTAATGMAAACATACAGCTGCATATAAACATAGA--TGCTGTGACAAATGCA 5651
Qy      6872 CATTAACCTGAGAACATACGCCATAGTAAGAACACATTAATATCA-----AGAGAACACA 6926
Db      5652 ATCTGAGCT--GATTATATGTCTATTAATTAACATGTAAGAGTCAGAGATTATAGTCTA 5709
Qy      6927 CAGCCATGTGTGGAGCCCATTTGGAGAGACACACAGCAAAAGTGAATGCAAAAAGAGAGA 6986
Db      5710 TAGTAAGTGTGTAGATCACTGGGGAGAACACAGACAAAGTAAGTGTGTGTGTGTGT 5769
Qy      6987 GAGAGTGAAGTGAAGATTGTGAAAAACAGGGCCACAGAAACACACAGAAATTAAGAGAG 7045
Db      5770 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5828

Search completed: December 15, 2003, 23:41:40
Job time : 23561.6 secs

```



XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-130909/17.  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 XX cytosine methylation -  
 PS Claim 1; SEQ ID NO 917; 32bp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 XX Sequence 5234 BP; 1479 A; 40 C; 1361 G; 2354 T; 0 other;  
 SQ  
 Query Match 15.1%; Score 1509.4; DB 24; Length 5234;  
 Best Local Similarity 79.9%; Pred. No. 3.3e-279;  
 Matches 1777; Conservative 0; Mismatch 446; Indels 0; Gaps 0;

QY 12 AGGAGGCCAGCAAAATCTGATGAGATCATGGGCTCTCAGAGGAGATTAAGGTTTCA 71  
 DB 3012 AGGAGGTTTATGAAATTTTGTAGATTTAGATGGGTTTGTGAAAGATTAAGTTG 3071  
 QY 72 TTGAGAAATTCCTGATGAGATCATGGGCTCTCAGAGGAGATTAAGGTTTCA 131  
 DB 3072 TTGAGAAATTTTGTAGAGATTAAGGCTTTTACAGAGGAGATTAAGTTTAC 3131  
 QY 132 AGGCTGAGAGGAAAGAAAGTGTAGGAGGAGTCTTAAATAGTGTCTCCATGCCAA 191  
 DB 3132 AGGTTGAGAGGAAAGAAAGTGTAGGAGGAGTCTTAAATAGTGTCTTAA 3191  
 QY 192 CACCCAGAAAGAACATGCTCCATGAGGAGAGATGATGAGATTTGGCTGA 251  
 DB 3192 TATTTAAAGAAAGATATGTTTGTATGAGGAGAGATGATGAGATTTGGCTGA 3251  
 QY 252 GCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311  
 DB 3252 GTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3311  
 QY 312 TATGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371  
 DB 3312 TATGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3371  
 QY 372 GCGTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 431  
 DB 3372 GTTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3431  
 QY 432 GTTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 491  
 DB 432 GTTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3491  
 QY 492 GGGATCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551  
 DB 492 GGGATCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3492  
 QY 552 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 611  
 DB 552 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3552  
 QY 612 TGACTCACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 671  
 DB 612 TGACTCACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3612  
 QY 672 ACTGAGATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 731  
 DB 672 ACTGAGATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3672

QY 732 AGAAGAGAGCTGATGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 791  
 DB 3732 AGAAGAGAGGATGATGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3791  
 QY 792 CATACCTACCCCAATTCCTATATCTCTACCTACCTACCTACCTACCTACCTACCTAC 851  
 DB 3792 TAT 3851  
 QY 852 TTGGCTGACACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 911  
 DB 3852 TTGGCTGAT 3911  
 QY 912 TTGACAAATATTCGGAACACCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 971  
 DB 3912 TTGACAAATATTCGGAACACCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3971  
 QY 972 ATATATATGATTAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1031  
 DB 3972 ATATATATGATTAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4031  
 QY 1032 GTAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1091  
 DB 4032 GTAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4091  
 QY 1092 ATGGCTAGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1151  
 DB 4092 ATGGCTAGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4151  
 QY 1152 AGCTAGAAATATGTAATTCAGTATGATGATGATGATGATGATGATGATGATGATGAT 1211  
 DB 4152 AGCTAGAAATATGTAATTCAGTATGATGATGATGATGATGATGATGATGATGATGAT 4211  
 QY 1212 GCTTTTCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1271  
 DB 4212 GCTTTTCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4271  
 QY 1272 TCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1331  
 DB 4272 TCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4331  
 QY 1332 CTTTTCAGAAAGCCCTTTCAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1391  
 DB 4332 CTTTTCAGAAAGCCCTTTCAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4391  
 QY 1392 TTGACACAGTGAACATTTATGCTCCAGACTCCTAGCCAGGAGGAGGAGGAGGAGGAG 1451  
 DB 4392 TTGACACAGTGAACATTTATGCTCCAGACTCCTAGCCAGGAGGAGGAGGAGGAGGAG 4451  
 QY 1452 GCGTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1511  
 DB 4452 GCGTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4511  
 QY 1512 TTTCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1571  
 DB 4512 TTTCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4571  
 QY 1572 GCTGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1631  
 DB 4572 GCTGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4631  
 QY 1632 CTCCCACTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1691  
 DB 4632 CTCCCACTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4691  
 QY 1692 TGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1751  
 DB 4692 TGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4751  
 QY 1752 TAGCCAGTATGCTTGAATTTGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1811  
 DB 4752 TAGCCAGTATGCTTGAATTTGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4811

QY	1812	GAGCTTCGATTCAGATTTCTGTCGACGCCGACCTAATTGCGAGTAGTAATCAGCC	1871
Db	4812	GAGCTTCGATTCAGATTTCTGTCGACGCCGACCTAATTGCGAGTAGTAATCAGCC	4871
QY	1872	AAATATGGCTTAAGTCTCGGAGACGACATTCGCCAGTAGAGTTGAGAGTGGGGGTGTG	1931
Db	4872	AAATATGGCTTAAGTCTCGGAGAGATAGTATATTTTATAGAGATTGAGAGTGGGGGTGTG	4931
QY	1932	CTGCTGCACACTATATATAGGAGAGTTCAACTGCTGCACCCAGAGCGTCGCTGCGCTCTG	1991
Db	4932	TTGTTGTTTAATTTATATATAGGAGATTATATGAGTTATTTAAGATGTTGTGTGTGTTTG	4991
QY	1992	CAGCTCAGCATGCTAGGGTACTGCGAGCACCCGTTGCACCTGGGTTGTGAGCCTATGC	2051
Db	4992	TAGTTTAGTATGTTAGGATTTAGGGATATTCGTTGATTTGGGTTGTGAGTTTAGT	5051
QY	2052	TGCTCTCTGCGCATTTGCCACCCCTCTCTCCGTCGATGAAGCTGGGACTGGAAGCCGAA	2111
Db	5052	TGCTTTTGTGATTTGTTATTTTTTTTTTTTTTCGTCGATGAAGCTGGGATTGGAAGCCGAA	5111
QY	2112	GATTGAGTTCTGGGCTAGGCTAGGATGAAGGCGCAGTTTTTAGGCCCTCGTCAAAATTTGGG	2171
Db	5112	GATTGAGTTTGGGCTTAGGCTTAGGCTTAGGCTTTAGTTTCCGTTAAATTTGGG	5171
QY	2172	TCAGGGGCTATGGAAGAGGATCGTCCCAATGCAATCAAGTATCTATTTGTCTCCCT	2231
Db	5172	TTAGGGGCTATGGAAGAGGATCGTCTTATATGATTAAGATATATTTGTTTTTTTT	5231
QY	2232	AGG 2234	
Db	5232	AGG 5234	

	RESULT 2
ABLJ32945/C	
ID ABLJ32945 standard; DNA; 5234 BP.	
AC ABLJ32945;	
DZ 26-MAR-2002 (first entry)	
DE Human immune system associated gene SEQ ID NO: 918.	
XX XX	
KW Human; immune system disease; cytosine methylation; antiasthmatic; KW antirheosclerotic; antiandemic; cytostatic; noctropic; KM neuroprotective; anti-HIV; anticoagulant; ophthalmological; KM antitumescetic; antiarthritis; antidabetic; antipsoriatic; KM antiflammatory; cancer; eye disease; arteriosclerosis; anaemia; KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; KM gene; dr.	
OS Homo sapiens.	
PQ WC00200928-A2.	
PN 03-JAN-2002.	
PD 02-JUL-2001; 2001WO-BP07537.	
PF 30-JUN-2000; 2000DE-1035529. PR 01-SEP-2000; 2000DE-1043826.	
PP (EPIG-) EPIGENOMICS AG.	
PA Olek A, Plegembrock C, Berlin K, PI WFI; 2002-130909/I7.	
DR Nucleic acid comprising fragment of chemically modified gene, useful PT for diagnosis and treatment of diseases associated with abnormal CYTOSINE METHYLATION - XX	

xx  
ps Claim 1, SEQ ID NO 918; 32bp + Sequence Listing; German.  
xx  
cc The present invention provides a number of human immune system associated  
cc genes which are modified by the methylation of cytosines. The sequences  
cc which are used in the diagnosis and treatment of immune system disorders,  
cc including eye diseases such as retinopathy, neovascular glaucoma and  
cc macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
cc leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
cc rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
cc diseases. The present sequence is a gene of the invention.  
xx  
sq Sequence 5234 BP; 1303 A; 40 C; 1091 G; 2800 T; 0 other;

Query Match	12.8%	Score 1281.6	DB 24	Length 5234
Best Local Similarity	73.4%	Pred. No. 1.3e-225		
Matches 1638	Conservative 0	Mismatches 594	Indels 0	Gaps 0
QY	1	TCCCTCTCCCGAGCAGGCGCCAGCAAAATCTGTAGATTCAGACAGGGTCTGCACAGCTG	60	
Db	2234	TCCTCTCCCGAAACAAACCCACAAATCTATTAATTCAAACAAATTTTAAACACTA	2175	
QY	61	AAGCAAGTTGTTAGAGAAATTCCTGATGAGATCAATGGGCTCTCAGAGGAGAAATA	120	
Db	2174	AAACAAATTTATTAATAAAATTTCTAATTAATAAAATCAATAATCTCAAAAAATAATA	2115	
QY	121	TAAGTTTCAGAGGCTGAGAGGAAAGAAAAGTGAGGGGAGTCTTGAATAGTGCTC	180	
Db	2114	TAAATTTCAAAAATTAATAAAAAATAAAAAATAAAAAATAAAATCTTAATAATTAATCTC	2055	
QY	181	CCATTGCCAACACCCAGAAAGAAAGACATGCGCCGATGGGAGGAAGTGATGATGAGA	240	
Db	2054	CCATTACCCAAACCCCAAAAAAAACAACTACCTCAATTAATAAAAAATAATTAATTAATA	1995	
QY	241	CATTGCGGTGAGCAGATGCGCATTTGCCAGGCTCGCAAGAGCTCAGAGATCCAGCCTT	300	
Db	1994	CATTACTATATACACAGTAAATCAATTACCAAACTACCAAAATCAAAAAATCCAACTT	1935	
QY	301	GCCCACTGACCTATGAGAGGGAATGATGTTCAACAACATTTTCATTGTAAGTCAGGA	360	
Db	1934	ACCCACTAACCTATTAATAAAAAATAATTAATTCACAACAACATTTTCATTGTAATTAATA	1875	
QY	361	GAGACACTGTGACCTGATGAGAGGCGCTGGTGACATGTTGTCAGAGGTTCCGGAATG	420	
Db	1874	AAAAACATTAACCTAATTAACAAAAACCTAATACTATTAATTCAAAAATTCGGAATA	1815	
QY	421	TGTGTTTTTCGTGTGAAGAACTTCGACAGATGAGAAAGGATCTGAGACTTTGGTA	480	
Db	1814	TATATTTTCCATTTAAAAAAAACCTTGGCAAAATTAATAAAAAATACTTAAATTAATA	1755	
QY	481	AGATTATATATGAGACTGTCAAGGGGTCTGAGCCATCTGTGAGGATCAAGGCCCTTCA	540	
Db	1754	AAATTAATATATTAATAAATCTATCAAAAATCTTAAAACCATCTATTAATAAATAAACCCTTCA	1695	
QY	541	GCTTGGCTAGGGAGCAGGGGCTCTGGAATTCATCTCGGCCCATAGTGAAGTCTGCCA	600	
Db	1694	ACCTTAATCTAAAAACAAAAATCTTAACCTTCACTCTTAACCCATACTAATCTAACCA	1635	
QY	601	TAATTTCTTTCTGACTACATGAGCAATCTCACAGAAAATGGGCGAGCTTTGGAGTGG	660	
Db	1634	TAATTTCTTTCTGACTACATGAGCAATCTCACAGAAAATGAACAACTTTAAAAATAA	1575	
QY	661	GCCAGAGAGTATCGAGGATGACAGTGAAGATCCACAGGAAGAGATGATGGGCGGAG	720	
Db	1574	ACCCAAAAAATCTAATAAATAAACAATTAATCCAAAAAATAAATAAATAAATAAACC	1515	
QY	721	ACACTGAGAGAGAAACAAGACTGTCAATTAAGGGCGTCTGTGACTCTAGATCTCAT	780	
Db	1514	ACACTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1455	
QY	781	ATGCTCTACTACATACTACCCCAATTCCTAATATTTCTCTACCTTAAGGGGGGGA	840	
Db	1454	ATACTCTACTACATACTACCCCAATTCCTAATATTTCTCTACCTTAATAAATAAATAA	1395	



QY	ATTGCGAAATTTGGCGCAACACTAGCAACACTAGCTAGCTTGGAAATGCAATTTTG	900
Db	1394 ATATCGAAAATTTACTACACACTTAACACACTACTCAACTTAATAATTCATTTTAA	1335
QY	CATTTTTCATTCACAAATATTTCTGAAACACTCTTATATGCGAGCACTATTTTATG	960
Db	1334 CATTTTTTCATTCAGAAATATTTCTTAAACACTCTTATATACCAACCTATTTTAA	1275
QY	961 GAGTCAGGGATATATATATGTTAAACAAGACAGGCAAAACAAGCAAGCAACACCA	1020
Db	1274 AAATCAAAAATATATTAATTAATTAACAAAACAACAAAACAACAAAACAACACCA	1215
QY	1021 TCACGAGTAATGTGACAGATGAAGAAATTTCAAGTTTATGTAAGTAAATTAACAAG	1086
Db	1214 TCACCAATTAATTAACAATTAACAAAATTTCAAAATTTTAATTAATTAATTAACAAAC	1155
QY	1081 AAGGCTCGAATGCGCTAGATTAAGCGGCTCAAGAAAGCTTCATTTGAGAGGTATGCT	1140
Db	1154 AAAATCTAAATATACCTAATTAATAATATCAAAAABAACTTCATTAABAAAATAATCT	1095
QY	1141 AAGCAGAGTCAAGTAAATATTTGTGAATTCAGTTACAGTTCTATTTGTTCTGGTT	1200
Db	1094 AAACAAAATCAACTAAAAATTAATTAATTAATTCATATTAATCTATTTATCTAAAT	1035
QY	1201 GGTTAATTAAGCTTTTCCCCCAAGGGAACCTAACAAAGAACTAATTAATCAATAG	1266
Db	1034 AATTAATTAATTAATCTTTTCCCCCAAAATTAACCAAAACAAAACCTAATTAATTA	975
QY	1261 TGTGTGTCTCTGGAAGAGAGACCTCTGTTTCTGCTCATCTACTGTCACCTTC	1320
Db	974 TAATATATCTCTCTAAABAAAACACCTCTCATTTCTACTCATTAATCAACCTTC	915
QY	1321 ACTTCAGGCACTTTTGCAAAAGCCCTTTCAGAGTCAGGGAAGCGAGGCTGGCATTG	1380
Db	914 ACTTCAAAACCTTTTCAAAAACCTTTTACCACATCAAAAAACGAAAACTAAACATA	855
QY	1381 GGGCTGGACATTTGACAAAGTGAATATTTGTCCCCGACCTCACTAGCCCAAGGT	1440
Db	854 AAACCTTAACCTTTAAACAACATTAATTAATTCACCAACCTCACTAACCCAAAT	795
QY	1441 AAAGCTGAAGAGCTTGGGCACTGCCCAAGAAAGCCCTGATGAAGCTTGGAAAAAGCTG	1500
Db	794 AAACTAABAAAACCTTAACATTAACCCCAAAAACCCCTATTAATTAACCTTAABAAAAC	735
QY	1501 TTCTCTGAGTATTTCTAAGTAAGTTATCTGTGTGTGTTACTTAAAGTAGTAAGTAT	1566
Db	734 TTCTCTAAATATTTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	675
QY	1561 TGTCTGTCTTAGCTGCTTAGAGCAGGGCTTGACACAGTACACAGCAATATTAATTCCT	1620
Db	674 TACTATCTCTAATCACTTAACCTTAAACAAACCTTAACACATTAATTAATTAATTCCT	615
QY	1621 CCTTTTCTCACTCCCACTATGTGGAGATAATCAATCAAAAAGGTATCCTCACTCT	1680
Db	614 CCTTTTCTCACTCCCACTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	555
QY	1681 ACTCACTTCCCTGACTTAATGAGTGTGAGCCATTTGCACTGTGAGAGTCAACCTGGA	1740
Db	554 ACTCACTTCCCTGACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	495
QY	1741 CGTGACAGAGTATAGCCCACTTAATCTGCTTGAATAATTTGCTGAAGGGGGTGGGGGCACTG	1800
Db	494 CGTCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	435
QY	1801 CCGGGAATAAGAGCTGTGATTAAGATTTCTGTCCAGACCTGACCTTAATTTGAGTGA	1866
Db	434 CCGAATAATAATAATTTTAATTTCAAAATTTCTATTCACAACTTAATTTTAATTAAT	375
QY	1861 TGTATCAAGCAATATTTGGCTTAAGTCTGTGAGACAGACATTTCCCACTAGAGTTGAGG	1920
Db	374 TATATATCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	315
QY	1921 TGGGGGTGTGTCTGTGCAACCTCTATATTAAGAGATTCACTGTGTCAACCAAGCTGTCC	1980

Db	Seq ID	Accession	Species	Gene	Length (bp)	GC (%)	GC (3bp)	GC (6bp)	GC (9bp)	GC (12bp)	GC (15bp)	GC (18bp)	GC (21bp)	GC (24bp)	GC (27bp)	GC (30bp)	GC (33bp)	GC (36bp)	GC (39bp)	GC (42bp)	GC (45bp)	GC (48bp)	GC (51bp)	GC (54bp)	GC (57bp)	GC (60bp)	GC (63bp)	GC (66bp)	GC (69bp)	GC (72bp)	GC (75bp)	GC (78bp)	GC (81bp)	GC (84bp)	GC (87bp)	GC (90bp)	GC (93bp)	GC (96bp)	GC (99bp)	GC (102bp)	GC (105bp)	GC (108bp)	GC (111bp)	GC (114bp)	GC (117bp)	GC (120bp)	GC (123bp)	GC (126bp)	GC (129bp)	GC (132bp)	GC (135bp)	GC (138bp)	GC (141bp)	GC (144bp)	GC (147bp)	GC (150bp)	GC (153bp)	GC (156bp)	GC (159bp)	GC (162bp)	GC (165bp)	GC (168bp)	GC (171bp)	GC (174bp)	GC (177bp)	GC (180bp)	GC (183bp)	GC (186bp)	GC (189bp)	GC (192bp)	GC (195bp)	GC (198bp)	GC (201bp)	GC (204bp)	GC (207bp)	GC (210bp)	GC (213bp)	GC (216bp)	GC (219bp)	GC (222bp)	GC (225bp)	GC (228bp)	GC (231bp)	GC (234bp)	GC (237bp)	GC (240bp)	GC (243bp)	GC (246bp)	GC (249bp)	GC (252bp)	GC (255bp)	GC (258bp)	GC (261bp)	GC (264bp)	GC (267bp)	GC (270bp)	GC (273bp)	GC (276bp)	GC (279bp)	GC (282bp)	GC (285bp)	GC (288bp)	GC (291bp)	GC (294bp)	GC (297bp)	GC (300bp)	GC (303bp)	GC (306bp)	GC (309bp)	GC (312bp)	GC (315bp)	GC (318bp)	GC (321bp)	GC (324bp)	GC (327bp)	GC (330bp)	GC (333bp)	GC (336bp)	GC (339bp)	GC (342bp)	GC (345bp)	GC (348bp)	GC (351bp)	GC (354bp)	GC (357bp)	GC (360bp)	GC (363bp)	GC (366bp)	GC (369bp)	GC (372bp)	GC (375bp)	GC (378bp)	GC (381bp)	GC (384bp)	GC (387bp)	GC (390bp)	GC (393bp)	GC (396bp)	GC (399bp)	GC (402bp)	GC (405bp)	GC (408bp)	GC (411bp)	GC (414bp)	GC (417bp)	GC (420bp)	GC (423bp)	GC (426bp)	GC (429bp)	GC (432bp)	GC (435bp)	GC (438bp)	GC (441bp)	GC (444bp)	GC (447bp)	GC (450bp)	GC (453bp)	GC (456bp)	GC (459bp)	GC (462bp)	GC (465bp)	GC (468bp)	GC (471bp)	GC (474bp)	GC (477bp)	GC (480bp)	GC (483bp)	GC (486bp)	GC (489bp)	GC (492bp)	GC (495bp)	GC (498bp)	GC (501bp)	GC (504bp)	GC (507bp)	GC (510bp)	GC (513bp)	GC (516bp)	GC (519bp)	GC (522bp)	GC (525bp)	GC (528bp)	GC (531bp)	GC (534bp)	GC (537bp)	GC (540bp)	GC (543bp)	GC (546bp)	GC (549bp)	GC (552bp)	GC (555bp)	GC (558bp)	GC (561bp)	GC (564bp)	GC (567bp)	GC (570bp)	GC (573bp)	GC (576bp)	GC (579bp)	GC (582bp)	GC (585bp)	GC (588bp)	GC (591bp)	GC (594bp)	GC (597bp)	GC (600bp)	GC (603bp)	GC (606bp)	GC (609bp)	GC (612bp)	GC (615bp)	GC (618bp)	GC (621bp)	GC (624bp)	GC (627bp)	GC (630bp)	GC (633bp)	GC (636bp)	GC (639bp)	GC (642bp)	GC (645bp)	GC (648bp)	GC (651bp)	GC (654bp)	GC (657bp)	GC (660bp)	GC (663bp)	GC (666bp)	GC (669bp)	GC (672bp)	GC (675bp)	GC (678bp)	GC (681bp)	GC (684bp)	GC (687bp)	GC (690bp)	GC (693bp)	GC (696bp)	GC (699bp)	GC (702bp)	GC (705bp)	GC (708bp)	GC (711bp)	GC (714bp)	GC (717bp)	GC (720bp)	GC (723bp)	GC (726bp)	GC (729bp)	GC (732bp)	GC (735bp)	GC (738bp)	GC (741bp)	GC (744bp)	GC (747bp)	GC (750bp)	GC (753bp)	GC (756bp)	GC (759bp)	GC (762bp)	GC (765bp)	GC (768bp)	GC (771bp)	GC (774bp)	GC (777bp)	GC (780bp)	GC (783bp)	GC (786bp)	GC (789bp)	GC (792bp)	GC (795bp)	GC (798bp)	GC (801bp)	GC (804bp)	GC (807bp)	GC (810bp)	GC (813bp)	GC (816bp)	GC (819bp)	GC (822bp)	GC (825bp)	GC (828bp)	GC (831bp)	GC (834bp)	GC (837bp)	GC (840bp)	GC (843bp)	GC (846bp)	GC (849bp)	GC (852bp)	GC (855bp)	GC (858bp)	GC (861bp)	GC (864bp)	GC (867bp)	GC (870bp)	GC (873bp)	GC (876bp)	GC (879bp)	GC (882bp)	GC (885bp)	GC (888bp)	GC (891bp)	GC (894bp)	GC (897bp)	GC (900bp)	GC (903bp)	GC (906bp)	GC (909bp)	GC (912bp)	GC (915bp)	GC (918bp)	GC (921bp)	GC (924bp)	GC (927bp)	GC (930bp)	GC (933bp)	GC (936bp)	GC (939bp)	GC (942bp)	GC (945bp)	GC (948bp)	GC (951bp)	GC (954bp)	GC (957bp)	GC (960bp)	GC (963bp)	GC (966bp)	GC (969bp)	GC (972bp)	GC (975bp)	GC (978bp)	GC (981bp)	GC (984bp)	GC (987bp)	GC (990bp)	GC (993bp)	GC (996bp)	GC (999bp)	GC (1002bp)	GC (1005bp)	GC (1008bp)	GC (1011bp)	GC (1014bp)	GC (1
----	--------	-----------	---------	------	-------------	--------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------	-------------	-------------	-------------	-------------	-------

	PR	14-AUG-2000;	2000US-0225758.
	PR	14-AUG-2000;	2000US-0225759.
	PR	18-AUG-2000;	2000US-0226279.
	PR	22-AUG-2000;	2000US-0226681.
	PR	22-AUG-2000;	2000US-0226686.
	PR	22-AUG-2000;	2000US-0227182.
	PR	23-AUG-2000;	2000US-0227009.
	PR	30-AUG-2000;	2000US-0228924.
	PR	01-SEP-2000;	2000US-0229287.
	PR	01-SEP-2000;	2000US-0229343.
	PR	01-SEP-2000;	2000US-0229344.
	PR	01-SEP-2000;	2000US-0229345.
	PR	05-SEP-2000;	2000US-0229599.
	PR	05-SEP-2000;	2000US-0229513.
	PR	06-SEP-2000;	2000US-0230437.
	PR	06-SEP-2000;	2000US-0230438.
	PR	08-SEP-2000;	2000US-0231242.
	PR	08-SEP-2000;	2000US-0231243.
	PR	08-SEP-2000;	2000US-0231244.
	PR	08-SEP-2000;	2000US-0231413.
	PR	08-SEP-2000;	2000US-0231414.
	PR	08-SEP-2000;	2000US-0232080.
	PR	12-SEP-2000;	2000US-0231968.
	PR	14-SEP-2000;	2000US-0232337.
	PR	14-SEP-2000;	2000US-0232338.
	PR	14-SEP-2000;	2000US-0232339.
	PR	14-SEP-2000;	2000US-0232401.
	PR	14-SEP-2000;	2000US-0232401.
	PR	14-SEP-2000;	2000US-0233063.
	PR	14-SEP-2000;	2000US-0233064.
	PR	14-SEP-2000;	2000US-0233065.
	PR	21-SEP-2000;	2000US-0234223.
	PR	21-SEP-2000;	2000US-0234274.
	PR	25-SEP-2000;	2000US-0234997.
	PR	25-SEP-2000;	2000US-0234998.
	PR	26-SEP-2000;	2000US-0235484.
	PR	27-SEP-2000;	2000US-0235834.
	PR	27-SEP-2000;	2000US-0235836.
	PR	29-SEP-2000;	2000US-0236327.
	PR	29-SEP-2000;	2000US-0236367.
	PR	29-SEP-2000;	2000US-0236368.
	PR	29-SEP-2000;	2000US-0236369.
	PR	29-SEP-2000;	2000US-0236370.
	PR	02-OCT-2000;	2000US-0236802.
	PR	02-OCT-2000;	2000US-0237037.
	PR	02-OCT-2000;	2000US-0237038.
	PR	02-OCT-2000;	2000US-0237039.
	PR	02-OCT-2000;	2000US-0237040.
	PR	13-OCT-2000;	2000US-0239935.
	PR	13-OCT-2000;	2000US-0239937.
	PR	20-OCT-2000;	2000US-0240960.
	PR	20-OCT-2000;	2000US-0241221.
	PR	20-OCT-2000;	2000US-0241785.
	PR	20-OCT-2000;	2000US-0241786.
	PR	20-OCT-2000;	2000US-0241787.
	PR	20-OCT-2000;	2000US-0241808.
	PR	20-OCT-2000;	2000US-0241809.
	PR	20-OCT-2000;	2000US-0241826.
	PR	01-NOV-2000;	2000US-0244617.
	PR	08-NOV-2000;	2000US-0246474.
	PR	08-NOV-2000;	2000US-0246475.
	PR	08-NOV-2000;	2000US-0246476.
	PR	08-NOV-2000;	2000US-0246477.
	PR	08-NOV-2000;	2000US-0246478.
	PR	08-NOV-2000;	2000US-0246523.
	PR	08-NOV-2000;	2000US-0246524.
	PR	08-NOV-2000;	2000US-0246525.
	PR	08-NOV-2000;	2000US-0246526.
	PR	08-NOV-2000;	2000US-0246527.
	PR	08-NOV-2000;	2000US-0246528.
	PR	08-NOV-2000;	2000US-0246529.
	PR	08-NOV-2000;	2000US-0246532.
	PR	14-AUG-2000;	2000US-0225758.
	PR	14-AUG-2000;	2000US-0225759.
	PR	18-AUG-2000;	2000US-0226279.
	PR	22-AUG-2000;	2000US-0226681.
	PR	22-AUG-2000;	2000US-0226686.
	PR	22-AUG-2000;	2000US-0227182.
	PR	23-AUG-2000;	2000US-0227009.
	PR	30-AUG-2000;	2000US-0228924.
	PR	01-SEP-2000;	2000US-0229287.
	PR	01-SEP-2000;	2000US-0229343.
	PR	01-SEP-2000;	2000US-0229344.
	PR	01-SEP-2000;	2000US-0229345.
	PR	05-SEP-2000;	2000US-0229599.
	PR	05-SEP-2000;	2000US-0229513.
	PR	06-SEP-2000;	2000US-0230437.
	PR	06-SEP-2000;	2000US-0230438.
	PR	08-SEP-2000;	2



RESULT 4  
AAS32492  
ID AAS32492 standard; DNA; 32191 BP.  
XX AAS32492;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 446.  
XX  
KW Human; endocrine antigen; ds; cytostatic; antifertility; antidiabetic;  
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;  
KW antiense-therapy; antibody; endocrine disorder; hormone imbalance;  
KW reproductive disorder; endocrine cancer; pancreatic disorder;  
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;  
KW hyperthyroidism; hypochatamic disorder; vanishing testes syndrome.  
XX  
OS Homo sapiens.  
XX  
PN MO200155319-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01335.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234323.  
PR 21-SEP-2000; 2000US-0234374.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.



QY 8474 ATGCACCAAGTAGTCCCGAGCTGTTCAAGAGGCTGAGGT -GGAGAGATTGCTCAGATGCGG 8532  
Db 4344 GCCTGCCGGTGTGCCAGTACTTGGAGGCTGAGGTGGAGATTGCTTGAAGTCAGA 4403  
QY 8533 AGGCAGAGATTGAGTGAACCGTACTGTGCTTCGCAATCCAGCCTGGGTGACAGATTG 8592  
Db 4404 CGGTCAAGCTTACAGTGAAGCTGATA-TGTGCCACTGCAAGTCCAGCCTGGATGACAAAGTG 4462  
QY 8593 AGACCTGTCTCAAAAAACAATTAATTAATAAATAA 8632  
Db 4463 AAACCTGTCTCAAAAGAAAAAAATGAGTATATCATA 4502

RESULT 5  
AAK68939/C  
ID AAK68939 standard, DNA; 28180 BP.  
XX AAK68939;  
XX  
AC  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23751.  
XX  
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; dc.  
OS Homo sapiens.  
XX  
XX W0200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JUN-2001; 2001MO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 14-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.







	PR	08-NOV-2000;	2000US-0246532.
	PR	08-NOV-2000;	2000US-0246509.
	PR	08-NOV-2000;	2000US-0246610.
	PR	08-NOV-2000;	2000US-0246611.
	PR	08-NOV-2000;	2000US-0246613.
	PR	17-NOV-2000;	2000US-0249207.
	PR	17-NOV-2000;	2000US-0249208.
	PR	17-NOV-2000;	2000US-0249209.
	PR	17-NOV-2000;	2000US-0249210.
	PR	17-NOV-2000;	2000US-0249211.
	PR	17-NOV-2000;	2000US-0249212.
	PR	17-NOV-2000;	2000US-0249213.
	PR	17-NOV-2000;	2000US-0249214.
	PR	17-NOV-2000;	2000US-0249215.
	PR	17-NOV-2000;	2000US-0249216.
	PR	17-NOV-2000;	2000US-0249217.
	PR	17-NOV-2000;	2000US-0249218.
	PR	17-NOV-2000;	2000US-0249219.
	PR	17-NOV-2000;	2000US-0249244.
	PR	17-NOV-2000;	2000US-0249245.
	PR	17-NOV-2000;	2000US-0249264.
	PR	17-NOV-2000;	2000US-0249265.
	PR	17-NOV-2000;	2000US-0249267.
	PR	17-NOV-2000;	2000US-0249269.
	PR	17-NOV-2000;	2000US-0249300.
	PR	01-DEC-2000;	2000US-0250391.
	PR	01-DEC-2000;	2000US-0251160.
	PR	05-DEC-2000;	2000US-0251030.
	PR	05-DEC-2000;	2000US-0251988.
	PR	05-DEC-2000;	2000US-0256719.
	PR	06-DEC-2000;	2000US-0251479.
	PR	08-DEC-2000;	2000US-0251856.
	PR	08-DEC-2000;	2000US-0251868.
	PR	08-DEC-2000;	2000US-0251869.
	PR	08-DEC-2000;	2000US-0251989.
	PR	08-DEC-2000;	2000US-0251990.
	PR	11-DEC-2000;	2000US-0254097.
	PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.		
PB			
PI	Rosen CA, Barash SC, Ruben SM;		
PK	WPI; 2001-541565/60.		
PL			
PM	Nucleic acids encoding 3224 human nervous system antigen polypeptides,		
PN	useful for preventing, diagnosing and/or treating nervous system		
PO	cancers and metastases -		
PP			
PQ	Disclosure; SEQ ID NO 13188; 1701bp + Sequence Listing; English.		
PS			
PT	The invention relates to novel genes (ABA11004-ABA21534) and proteins		
PU	(ABBA1678-ABBA18001) useful for preventing, treating or ameliorating		
PV	medical conditions e.g. by protein or gene therapy. The genes are		
PW	isolated from a range of human tissues disclosed in the specification.		
PX	The nucleic acids, proteins, antibodies and (ant)agonists are useful		
PY	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast		
PZ	and ovarian cancer and other cancers of the adrenal gland, bone, bone		
QA	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;		
QB	(b) immune disorders e.g. Addison's disease, allergies, autoimmune		
QC	hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's		
QD	disease, multiple sclerosis, rheumatoid arthritis and ulcerative		
QE	colitis; (c) cardiovascular disorders such as myocardial ischaemias;		
QF	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and		
QG	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal		
QH	and parasitic infections.		
QI	Note: The sequence data for this patent did not form part of the		
QJ	printed specification, but was obtained in electronic format directly		
QK	from WIPO.int/ftp.wipo.int/pub/published_pct_sequences.		
QL			
QM	Sequence 11234 BP; 3094 A; 2417 C; 2869 G; 2854 T; 0 other;		
QN			
QO			
QP	Query Match	3.6%; Score 357; DB 22; Length 11234;	
QQ	Best Local Similarity	53.5%; Fred. No. 1.2e-58;	

[illegible]

QY 8172 TCTGAATGCCCGCCATGATTTTTCAGAGCCACCATTAAGTGGTGTATGCCAGCC 8231  
 DB 3935 TATGTGTCACTAATAAATCCACAAACAAAGCAAGAGTAGGCCACCATTA 3876  
 QY 8232 TATGATATGGCTACATTAATGGGTGGGAATCATCATTTACTTCAATGCAAGAGC 8291  
 DB 3875 TGAGAAATGACAGAAATCATTAATAGAGAAATATGCCACAGAGACTTGGATATTGA 3816  
 QY 8292 ATAGCTCTGTTTCAAAATAGGCCCCCTGGCCAGGTGTGTGCTCATGCTATTATC 8351  
 DB 3815 ATTATATGTGTGATTAAGAAATATGTTTGCAGAGAGCGGTGGCTCAACCTGTATC 3756  
 QY 8352 CCAACATGAGGAGGCC-GAGGGGGGAGATCACTTGAAGTTCAGAGGCTTGAACAGC 8410  
 DB 3755 CAGCACTGTGGAGAACCGAGGTGGGTGATCACTGAGGTCAAGAGTGGAGACAGC 3696  
 QY 8411 CTGGGCAACATGATGAATCTCATCTCTAATAAATCAAAAAATAGCCAGGTGTG 8470  
 DB 3695 CTGGCCAAATATGTTGAATCCCGTGTCCACAAAATAC-AAAAATTAGCCGGCGTGT 3637  
 QY 8471 GCGATGACCATGATGTCCTCCAGCTGTTCAGAGGCTGAGGAGATTCCTGAGTGTG 8530  
 DB 3636 AGCAGCTGCTGTATGTCCTGGCTACCTGGAGGCTGAGGAGATCACTTGAACCCG 3577  
 QY 8531 GGAGGAGAGATTGCACTGTAACCTGTGCTGCAATGCAAGCTGGGTGACAGAT 8590  
 DB 3576 GGAGGGAGAGTGTGCACTGTAACAGATCACTGCACTGCACTGAGTGGGAGACAG 3517  
 QY 8591 TGAGACCTGTCTCAAAAAAATAAATAAATAAATATATGTTCTGAGCAGG 8650  
 DB 3516 CAAGATCTCATCTCAAAAAAATAAATAAATAAATATGTTA----- 3475  
 QY 8651 TAATTTCAATGGGAACTCTCCAGGGAGGTGATATGTCACCGCTGATACTCAG 8710  
 DB 3474 -----ATATATATAAGCCCTGATTCAGACCG 3446  
 QY 8711 TACAGGCTAATAAGAACTTGTGTAGCAGAGAAACCTGATTTACTCAACAA 8770  
 DB 3445 GACAGAAAATAAT 3386  
 QY 8771 TATTTGTGACATCTGATTAAGAGTGGGCAATGTCTTAGCAGTATACAGTATCA 8830  
 DB 3385 CCAATGCTGTGTTAAAAA-----AAAAA----- 3348  
 QY 8831 ACATGCGAACAAGATGCTGCTGCAAGGCTCTGCTAAGTGAAGAGACAATAAGA 8890  
 DB 3347 ACAGGCCAGGTGGGTGCTCAC-----ACCTGTAATCCGGCATTG 3303  
 QY 8891 AAGAGAAAGAGAAAGAAATATTTAGTAAATTAAGGTTGTAAGAAATAAGA 8950  
 DB 3302 GAGGGCAAGGACAGATCAGAGGTGAGAGATTTGAGAGTGGCTGGCCACATGTGTG 3243  
 QY 8951 CAGGATAGTGGATTAAGGTGAGAGAAATGAGGAGTGTCTTGAAGAAATGATTTTGA 9010  
 DB 3242 AAACCCCGTCTACTAATAAATACAAAAATCACTGAGGCTGATGAGCACTGCTGAAT 3183  
 QY 9011 GCTGAGACTTCAATGATGAGAGAAATTAACAACAGATGTGTGAGAGAAAGCATTTT 9070  
 DB 3182 CCCAGCTACTCGGAGGCTGAGGAGAGAGCTGTTGAATCGGGAGCA----- 3132  
 QY 9071 AAGGAGGATGAGAGACATTAATCTCAAGAAATCAAGAGAAACCTGTGTAGGCTGAAC 9130  
 DB 3131 ---GAGGTGACGTAGCGAGATCAAGCATTAATTAAGCTGGGTGACAGAACAG 3075  
 QY 9131 ACAGAGAAAGAGAGGTGGTGAATTAAGGAGGAGGAGAGGAGTGGCCAGGTTACTTA 9190  
 DB 3074 AATCTGACAAAAAAGATTTTAAAGAA-----CAATTTCTAGAA 3028  
 QY 9191 GACCTGTAAAGGTTTCAACATTAAGGAGTCAACAGAAAGTCTTGAAGAGGCTGT 9250  
 DB 3027 ATGTAGATACAGTAATTAACAATAAATAAATAAATAAATAAATAAATAAATAAATA 2968

QY 9251 GATATATCTAATCAATTTTAAAGATCACTGACTTGTTCAGAAATAGTTA 9310  
 DB 2967 GTGTGATGTGTAT 2908  
 QY 9311 TAAAGTACAGCATGTAAGACAGAAATCAGCTAGCAATCCGTGAGTGTCAATTA 9370  
 DB 2907 ACATATCAGAGATGTAGGAGTGAAT-----TTGCTCCCTTCTCGTACCAAAAT 2852  
 QY 9371 GAGGTATGACCGCTTGAAGTATGACAGCAGAGTGTGAGAAATCAATGAT 9430  
 DB 2851 CCTGTGTAAAGTCTTAACCCAGT-----ACCTGAGCTGGAT 2811  
 QY 9431 ATATTTGAGAGGAGTGAAGTGAAGATTAATAGTAAAGTAAAGGAGTGTG 9490  
 DB 2810 TGATTTGAGAAAGG-----TCTTAAAGATTAATAGTGTGCGGAGCAGTGTG 2756  
 QY 9491 CTACGCTGTATCTTACACTTTGGAGGCGCAAGGAGTGAATCAGTGAAGTCAAG 9550  
 DB 2755 CTGAGCTGTATCTTACAGACTTTGGAGGCTGAGACTGTGTGATCACTGAGTCAAG 2696  
 QY 9551 AGTTGAGACAGCTTGAACCAATGTGTGAACCTGTCTTAAATATAC-AAAAAT 9609  
 DB 2695 AGTTTGAAGACAGCTTGGCCACATGTGTGAACCTGTATCTTAAATAAATAAAT 2636  
 QY 9610 AGCTGGAAATGGTGGACATGCTGTATCTGAGCTTGTGGAGGCTGAGCAGAGA 9669  
 DB 2635 AGCCAGGCAATGGGGGAGTGTGCTGTATCTCAG-CTACTTGGAGGCTGAGCAGAGA 2577  
 QY 9670 ATGCTTGAACCTGGAGGTGAATGTTGCAAGTGAAGCCGAGATTGCAATTCAGTCCAG 9729  
 DB 2576 ATTGCTTGAACCTGGAGGAGAGAGTGTGCAAGTGAAGTTCGACATTCAGTCCAG 2517  
 QY 9730 CTTGGGAAACAGAGTGAACCTCGTCTTAAATAAATAAATAAATAAATAAATAAATA 9778  
 DB 2516 CTTGACTTAACAGAGTGAACCTCATCTCAAAAAAATAAATAAATAAATAAATAAATA 2468

RESULT 7  
 ABK22784  
 ID ABK22784 standard; cDNA; 57273 BP.  
 XX  
 AC ABK22784;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Human high bone mass (HBM) polynucleotide clone #7.  
 XX  
 KW Human; mouse; Zmax1; HBM; high bone mass gene; lipid regulation; stroke;  
 KW lipid-associated condition; arteriosclerosis; cardiovascular disease; B3;  
 KW osteoporosis; atherosclerosis; diabetic atherosclerosis; plaque build-up;  
 KW neurovascular condition; wound healing; gene therapy; PCR primer; probe;  
 KW bone development disorder; antiarteriosclerotic; cardiovascular;  
 KW osteopathic; cerebroprotective.  
 XX  
 OS Homo sapiens.  
 PN WO200192891-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PP 25-MAY-2001; 2001WO-US16946.  
 XX  
 PR 26-MAY-2000; 2000US-0578900.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 PA (UYCR-) UNIV CREIGHTON SCHOOL MEDICINE.  
 XX  
 PI Carulli JP, Little RD, Recker RR, Johnson ML;  
 XX  
 DR WPI; 2002-097784/13.  
 XX  
 PT Identifying molecules involved in lipid regulation, useful for  
 diagnosing, treating or preventing e.g., arteriosclerosis, comprises



XX WO200177327-A1.  
 XX 18-OCT-2001.  
 XX 21-JUN-2000; 2000MO-US16951.  
 XX 05-APR-2000; 2000US-0543771.  
 XX 05-APR-2000; 2000US-0544398.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX Carulli JP, Little RD, Recker RR, Johnson ML;  
 XX MPI; 2001-657171/75.  
 XX  
 XX New high bone mass (HBM) and Zmaxi genes and proteins useful for  
 XX modulating bone mass for the treatment of e.g. osteoporosis -  
 XX  
 XX Claim 51; Page 308-350; 443pp; English.  
 XX  
 XX The present invention describes the human Zmaxi gene and the high bone  
 XX mass (HBM) gene, which are found on chromosome 11q13.3. The Zmaxi and  
 XX HBM genes have osteopathic activities. The genes can be used in gene  
 XX therapy, antisense therapy and in the production of vaccines. They  
 XX can be used in the diagnosis and treatment of bone disorders including  
 XX osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous  
 XX dysplasia. ABA82038 to ABA82700 and ABA68168 to ABA68193 represent  
 XX sequences used in the exemplification of the present invention.  
 XX  
 XX Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;  
 SQ  
 Query Match 3.5%; Score 355; DB 22; Length 66933;  
 Best Local Similarity 59.6%; Pred. No. 4.1e-58;  
 Matches 871; Conservative 0; Mismatches 460; Indels 130; Gaps 11;  
 QY 7184 TTAATAATACCCAGCGCGGCGTGTGAGTCAAGCCTGTAATCTAGCACTTTGGAGGC 7243  
 Db 61260 TTAAGAATCTTGGGCGCCAGCGTGTGAGTCAAGCCTGTAATCTAGCACTTTGGAGGC 61319  
 QY 7244 TGAAGAGGTGATGCTTGTAGCGGCTTGAAGCTTGAAGTTTGAACCAAGCTTGGCAAC 7303  
 Db 61320 TGAAGAGGTGATGCTTGTAGCGGCTTGAAGCTTGAAGTTTGAACCAAGCTTGGCAAC 61370  
 QY 7304 ATGGCAAACTATCTTCAAAAATATAAATTAATGTCGGGTGTGATGCTTCC 7363  
 Db 61371 ATGGCAAACTATCTTCAAAAATATAAATTAATGTCGGGTGTGATGCTTCC 61430  
 QY 7364 TATAG-----TCCATCTACTTCAAGCGCTGAGCGGAAAGTCAAGGCTTCAAGTACC 7418  
 Db 61431 ACTTGGAGTCTCAAGCAATGAGATTTGCTTGAACATGAGGCGAGGTTGCAAGTACC 61490  
 QY 7419 GTGATGCTGCTACTGCACTTCAAGCTTGTGAGAGTGAAGCAATGCTCAAAAAAAC 7478  
 Db 61491 GAGATGCGGCTATGCACTTCAAGCTTGTGAGAGTGAAGCAATGCTCAAAAAAAC 61550  
 QY 7479 AAAAAAAGAAAGAAAGAAAGAAATTAATCAATCTCTTTTATGGAATATAATATAT 7538  
 Db 61551 AAAAAAAGAAAGAAAGAAAGAAATTAATCAATCTCTTTTATGGAATATAATATAT 61610  
 QY 7539 ACTAGCTGAGTGAAGTCAATTTTGTGTAATTAATCAACCAATGACCCACTACAGC 7598  
 Db 61611 AT 61663  
 QY 7599 ATGGGAGAGTCCCTCCCTCTCACTGTAATTTTCTTCTGACTCAAGTTTGT 7658  
 Db 61664 ---AGGCTTCCCTGCTGATCATATAAACAATATTTTCACTCTCTCTTTT 61720  
 QY 7659 GTTGTGCTGCTGCTTGTGATGAGTCACT-CTGTACCCAGGCTGAGTGCAT 7717  
 Db 61721 TT-----TTTGAGACAGAGTTTGTCTCTGTGCTGAGGCTGAGTGCAGT 61766  
 QY 7718 GGGCAATCTGTGTTCACTGCAACTCTGCTCTGGGTTCAAGCAATCTCTGCTCA 7777

Db 61767 GGGCAATCTGAGTCACTGTAACCTCCGCTCCGGGTGAGATATCTCTGCTCA 61826  
 QY 7778 GCTTCCCGATATAGTGGAGCTACAGGCGATACACCAATGCTGCTGCTAATTTTGTATTT 7837  
 Db 61827 CTTTCCCGATATAGTGGAGCTTATAGCAATGACACCAATGCTGCTGCTAATTTTGTATTT 61875  
 QY 7838 TTTTGGGTTTCAATGATATATTTATTTATTTTGTATTTTGTATTTTGTATTTTGTATTT 7896  
 Db 61876 -----TTTGTATCTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 61899  
 QY 7897 GGTTCACATGTTGGCCAGGCTGCTGAACTCCCTGACTGAGTGAATCCGCTGCT 7956  
 Db 61900 GGTTCACATGTTGGCCAGGCTGCTGAACTCCCTGACTGAGTGAATCCGCTGCT 61959  
 QY 7957 CGGCTCCCAAGTGTAGATTTACAGGATGACACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 8016  
 Db 61960 CGGCTCCCAAGTGTAGATTTACAGGATGACACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 62019  
 QY 8017 TTTATCTATCTCTTCTCTTCTTCAACTTTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 8076  
 Db 62020 TCTTAAGTCT 62058  
 QY 8077 CCACTTCTCACTACTTCACTGATCTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 8136  
 Db 62059 TAACTTCTGAGATGATTTATGATTTATGATTTATGATTTATGATTTATGATTTATGATTTAT 62118  
 QY 8137 GCTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 8196  
 Db 62119 GCTGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 62178  
 QY 8197 AGACCAACACATTTAATGATGATTTGCTGATGATTTGCTGATGATTTGCTGATGATTTGCTGAT 8256  
 Db 62179 AG-----AATTTGAACAGAGAAAGTTCCTGATCAAGGCTTATTTACAGGATGAGAA 62230  
 QY 8257 GGGAACTCATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 8316  
 Db 62231 TACAGAAATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62277  
 QY 8317 CCCTGGCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8376  
 Db 62278 GAATAGGCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62337  
 QY 8377 GCAATCACTTGAAGTCCAGGATTTCTAGACCAAGCTTGGGCAACATGCTGAATCTATCT 8436  
 Db 62338 GTGATCACTTGAAGTCCAGGATTTCTAGACCAAGCTTGGGCAACATGCTGAATCTATCT 62397  
 QY 8437 CTACTAAATAATAC-AAAAAATTAGCAAGTGTGCTGATGCAAGCAAGTGTGCTGATGCTGATGCTGAT 8495  
 Db 62398 CTACTAAATAATACAAAAAATTAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 62457  
 QY 8496 TCAGAGGCTGAGTGGAGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 8555  
 Db 62458 TCGGAGGCTGAGGAGGAGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 62517  
 QY 8556 GACTGCTCTGCTGATCACTGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8615  
 Db 62518 GATCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 62577  
 QY 8616 AATATAATATAATAATATG 8636  
 Db 62578 AATAACAGAGTTCACTTG 62598  
 Db 62598 AATAACAGAGTTCACTTG 62598  
 RESULT 9  
 ACC45366  
 ID ACC45366 standard; DNA; 66933 BP.  
 XX ACC45366;  
 AC  
 XX 02-JUN-2003 (first entry)  
 DT  
 XX  
 XX Human HBM gene fragment #7.

XX Human; high bone mass; HBM; LRP5; LRP6; transgenic; bone mass modulation;  
 KW gene therapy; bone density modulation; bone strength; trabecular number;  
 KW bone size; bone tissue connectivity; bone disease; osteoporosis;  
 KW osteomalacia; rickets; Paget's disease; neoplasm of the bone; gene; ds.  
 OS Homo sapiens.  
 XX MO200292764-A2.  
 XX 21-NOV-2002.  
 XX 13-MAY-2002; 2002MO-US14876.  
 XX 11-MAY-2001; 2001US-290071P.  
 XX 17-MAY-2001; 2001US-291311P.  
 PR 01-FEB-2002; 2002US-353058P.  
 PR 04-MAR-2002; 2002US-361293P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA (AMHP) MYETH.  
 XX Babit J, Bex FJ, Yaworsky PJ, Bodine PV;  
 PI WPI; 2003-129278/12.  
 DR New transgenic animals (e.g. mice), useful as models for studying bone  
 XX density modulation, developing drugs for treating or preventing bone  
 PT diseases (e.g. osteoporosis), or diagnosing diseases characterized by  
 PT reduced bone density -  
 XX Example 2; Page 362-384; 603pp; English.  
 XX The invention relates to novel transgenic animals expressing the high  
 CC bone mass (HBM) gene, expressing the corresponding wild type HBM gene,  
 CC comprising an alteration of the gene encoding LRP5 or LRP6, or  
 CC expressing an LRP5 that is modulated by an altered gene control  
 CC sequence introduced by homologous or non-homologous recombination. The  
 CC transgenic animals are for the study of bone density modulation or bone  
 CC mass modulation. The invention has osteopathic and cytostatic activity.  
 CC The polynucleotides of the invention may have a use in gene therapy.  
 CC The transgenic animals and nucleic acids are for the study of  
 CC bone density modulation, where the bone mass is modulated relative to  
 CC non-transgenic animals of the same species in more than one parameter  
 CC selected from bone density, bone strength, trabecular number, bone  
 CC size, or bone tissue connectivity. The transgenic animals, nucleic  
 CC acids and methods are useful for identifying molecules involved in bone  
 CC development, and for developing pharmaceutical compositions, which may  
 CC be employed for treating or preventing bone diseases, e.g.  
 CC osteoporosis, osteomalacia, rickets, Paget's disease, or neoplasms of  
 CC the bone. The transgenic animals and nucleic acids are also useful in  
 CC methods for diagnosing diseases involved in bone development, or  
 CC characterised by reduced bone density or mass. The present sequence is  
 CC used in the exemplification of the invention.  
 XX Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;  
 SQ  
 Query Match 3.5%; Score 355; DB 25; Length 66933;  
 Best Local Similarity 59.6%; Pred. No. 4.1e-58;  
 Matches 871; Conservative 0; Mismatches 460; Indels 130; Gaps 11;

QY 7364 TATAG-----TCCCATCTACCTGAGGCTGAGCCCGGAAGGTCAGGCTTCAGTAGACC 7418  
 Db 61431 ACTTGGAGCTGATAGACATGAGTAATGCTTGAACATGAGGAGGAGGTTGACAGTGGCC 61490  
 QY 7419 GTGATGCTGCTACTGCACTCCAGCTGGGTGACAGAGTGAACATGCTCCAAAAAAC 7478  
 Db 61491 GAGATGCGCCGATGTGACCTCCAGCTGGGCGACAGAGCGAGATCTGCTCAAAAAAAA 61550  
 QY 7479 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 7538  
 Db 61551 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 61610  
 QY 7539 ACCTAGCTGAGTGAAGATGACATATTTGGTAAATACCAACATGACCACTACAGC 7598  
 Db 61611 AT 61663  
 QY 7599 ATGGGCACTGCTCCCT 7658  
 Db 61664 ---AGGCTTCCCTGCTCTGATCATAAAAAACAATATTTTCACTCTCTCTCTCTCTCT 61720  
 QY 7659 GTTGTGTTGCTGTTGTTGAGATGAGTCTCACT-CTGTCAACCAAGCTGAGTGCAT 7717  
 Db 61721 TT-----TTTGAACAAGATTTTGTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 61766  
 QY 7718 GGGCAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7777  
 Db 61767 GGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61826  
 QY 7778 GCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7837  
 Db 61827 CTTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61875  
 QY 7838 TTTTGGGTTAACAATGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7896  
 Db 61876 -----TTTGTACTTTAGTAGACGG 61899  
 QY 7897 GATTTCACATGTTGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7956  
 Db 61900 GATTTCACATGTTGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61959  
 QY 7957 CGGCTTCCCAATGCTGATGATTTACAGGATGAGCAACAGCTGCTGCTGCTGCTGCTGCTG 8016  
 Db 61960 CGGCTTCCCAATGCTGATGATTTACAGGATGAGCAACAGCTGCTGCTGCTGCTGCTG 62019  
 QY 8017 TTTATGATCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8076  
 Db 62020 TCTTAACTGCT 62058  
 QY 8077 CCACTTCTCACTTCACTGATGATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8136  
 Db 62059 TAAATCTTCTGAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 62118  
 QY 8137 GCTCTTTTCTCTGATGATTTCCAGAGAGGAATCTGAATGCCAGTCAATTTTC 8196  
 Db 62119 GCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 62178  
 QY 8197 AGACACACACATTAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8256  
 Db 62179 AG-----AATTGAACAGAAAGTTCCTGCTGACAGCTTATTAACAGGACCTGGA 62230  
 QY 8257 GGGAACTCATCTTATCTTCACTGCAACAAAGCAATAGCTGCTGCTGCTGCTGCTGCTGCTG 8316  
 Db 62231 TACAGAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 62277  
 QY 8317 CCGTGGGCGAGGT 8376  
 Db 62278 GAAATAGCCAGGATGT 62337  
 QY 8377 GCAATCACTTGTGCTGAGAGTGTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8436  
 Db 62338 GTGATCTACTGAGTGTGAGAGTGTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62397  
 QY 8437 CTACTAAAAATAC-AAAAAATTAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8495

Db 62398 CTAATAAATAACAAAAAATTAGCTGGTGGTGGCGATGCTGTATCCAGCTAC 62457  
QY 8496 TCAGAGGCTGAGCTGGAGAGATTGCTCGAGTGTGGAGCGAAGATTGCAATGAAACCGT 8555  
Db 62458 TCGGAGGCTGAGCGAGGAGAAATCACTTGAACCTGGAGCGAGGTTGCAATGAGCCGA 62517  
QY 8556 GACTGTCCTCTGCAATCCAGCTGGGTGACAGATTGAGACCCGTCTCAAAAAACAAT 8615  
Db 62518 GATCATGCCCTTACTCTCCAGCTGGTGGAGAGGCCGAATCTGTGAAAAA 62577  
QY 8616 AATTAATTAATTAATTAATG 8636  
Db 62578 AAAAAAGAAAGTTCAACTTG 62598  
  
RESULT 10  
AAL05708  
ID AAL05708 standard; DNA; 27154 BP.  
XX AAL05708;  
AC  
XX 21-NOV-2001 (first entry)  
DT  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 8396.  
DS  
XX Human; reproductive system related antigen; reproductive system disorder;  
KM cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
PD  
XX 02-AUG-2001.  
PF  
XX 17-JAN-2001; 2001WO-US01339.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234423.  
PR 21-SEP-2000; 2000US-0234424.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.







Oy	7640	TCTGACTCAGAGTTTGTGTTGTGTTGTGCTGTGTTGTGTAATGAGTCACTACCTGCA	7699
Db	7906	TATTATATATATGTATTATTAATTTT-----TTTGAGAGGACCTCTGCTCTGCG	7957
Oy	7700	CCAGAGCTGAGTGCATAGGCGCAATCTTGATTCACTGCAACCTCTGCTCTGGATTCA	7759
Db	7958	TCCAGGCTGAGAGCCGAGTGGCGGCATCTAGCTCCACCAACTCTCACTCTGGATTCA	8017
Oy	7760	AGGCAATCTCTGCTGCTGAGCCTCCCGTATAGCTGGGAATTAAGGCGCATACCACTAGCC	7819
Db	8018	AGCAATCTCTGCTGCTGAGCCTCCG---ACTGGAGCTAAGGACGACCATCATATC	8073
Oy	7820	TGGCTAATTTTGTATTTTGTGTTTGGTTTACATATGATCTATTTATTAATTTTGT	7879
Db	8074	TGGCTAGTTTGTATTTTGTATTTTGTAGT---AAGCGGATTTTGCCATGTGGCCAGGCTG	8129
Oy	7880	ATTTTGTATGATAGATAGGATTGATTCACCAATGTGGCCAGGCTGATCTCGAATCTGACCTC	7939
Db	8130	TCTCGAATCTCTGACTCAGTGAATCATTCTGCTCCCAAGTGTGGATTACAGGCGTG	8189
Oy	7940	AGGTGATCCGCGCTGCTGCTGCGCTCCCAAGTGTAGATTTACAGGCATGACCAACCGC	7999
Db	8190	AGCAACCGTGGCCGGCCGATTTCCCATTTTATTAAGGCCACAG--TCAATGGGATTA	8247
Oy	8000	CTGGCCCCCTCAAGGTTTATTTATCTATCTCTTGGCTTTTGCACAACCTTGGCTTGACGT	8059
Db	8248	AGGGCCCAACCGTGGCTCAGATGACCTCATCTTAACATATGCAATCTGCCAACAACATTT	8307
Oy	8060	GGACCATATGTTCTCTCCACTTCTCACTTCTCATCTTCACTTCAAGTCTCAAGTTCCAA	8119
Db	8308	TCCAAATTAACATCTCATGCTGAGGTCTGGGGGTTAGGATGTATCATATGATTTTGAT	8367
Oy	8120	TGATACCTCCCTCAGTTGCTCTTTTCTCTATGTAAGATTTCCAGAGGGAACTGATG	8179
Db	8368	GGGACAGTTCTTTATTTTCTTTTATTTTAAAGACAGAGGCTTGCTCTGTACCCAGGC	8427
Oy	8180	GCCCAATCCATATTTTACAGCACACCAATTAAGTGTGATTTGCCAGGCTATGATTT	8239
Db	8428	TGAGTGAAGTGTGTATCATATGCTACCTGACGCTGATCTTGGGCTCAAGTATC	8487
Oy	8240	GGCTACATTAATGGTGTGGAACTCATATTTACTTCAATGCAAAAGACATAGCTCT	8299
Db	8488	CTTCTGCTGGGTTTCCAGTATGTCGAATTAACAGGTGACACACCACTGGCTTAA	8547
Oy	8300	GGTCTCTCAAAATAG--GCCCTGGGGCAGGTGTGGTGCCTCATGCTTAATCCCA	8357
Db	8548	TTTTTATATTTTGTATGAGACAGGGGCTGGGGTGTGGCTCAACCCCTGTAATCTCACA	8607
Oy	8358	CTGTGGAGGCCGAGGGGGGAGATCACTTAGTTCAGAGATTTAGACCAAGCTGGGCA	8417
Db	8608	CTTTGGGAGGCCGAGGCGAGTGTATCATCTTAAGTCAAGGATTTGAGACCAAGCTGGGCA	8667
Oy	8418	ACATGTGTAAATCTTCATCTCTACTATAAAATACAAAAATTGACGAGGTGTGTGATGC	8477
Db	8668	ACATGTGTAAATCCCGCTCTCTACTATAAAATAC--AAATTTGACGAGGTGTGTGTGAT	8725
Oy	8478	ACCAATGATCCAGCTGTTCAGAGAGGCTGAGGTGGAGAGATTGCTCGAGTGTGGAGGCA	8537
Db	8726	GCTGTATATCCAGTACTTGTGGAGGCTGAGGCGAGAAATTTGCTTGAACCCGGGAGGCC	8785
Oy	8538	GAGATTGCAATGAACCGTGACTGTGCTCTGCAATCCAGCTGGGTGACAGATTGAGACC	8597
Db	8786	GAGGTGCAATGAGCTAGGATTCTGTGCCACTGCACTCCAAATCTGGGTGACAGCAAGCT	8845
Oy	8598	CTGTCTCAAAAAACAATTAATTAATTAATTAATA	8634
Db	8846	CCGTCTCAAAAAAAAAAAAAAAAAAGAGAGA	8882

RESULT 12  
AAS32516  
ID AAS32516 standard; DNA; 29543 BP  
XX

AC	AAS32516;	
XX		
DT	17-DEC-2001	(first entry)
DE	Human genomic DNA for novel endocrine antigen, SEQ ID No 470.	
XX		
KW	Human; endocrine antigen; des; cytostratic; antifertility; antidiabetic;	
KW	thyroid-active; adrenal-active; androgenic; gastric; gene therapy;	
KW	antisenase-therapy; antibody; endocrine disorder; hormone imbalance;	
KW	reproductive disorder; endocrine cancer; pancreatic disorder;	
KW	diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;	
XX	hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.	
OS	Homo sapiens.	
XX		
PN	WO200155319-A2.	
XX		
PD	02-AUG-2001.	
XX		
PE	17-JAN-2001; 2001WO-US01335.	
XX		
PR	31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000;	2000US-0180628.
PR	24-FEB-2000;	2000US-0184664.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0189874.
PR	17-MAR-2000;	2000US-0190076.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	07-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	07-JUL-2000;	2000US-0216880.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217496.
PR	14-JUL-2000;	2000US-0218290.
PR	26-JUL-2000;	2000US-0220963.
PR	26-JUL-2000;	2000US-0220964.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.
PR	14-AUG-2000;	2000US-0225213.
PR	14-AUG-2000;	2000US-0225214.
PR	14-AUG-2000;	2000US-0225286.
PR	14-AUG-2000;	2000US-0225287.
PR	14-AUG-2000;	2000US-0225288.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225757.
PR	14-AUG-2000;	2000US-0225758.
PR	14-AUG-2000;	2000US-0225759.
PR	18-AUG-2000;	2000US-0226279.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226688.
PR	22-AUG-2000;	2000US-0227182.
PR	23-AUG-2000;	2000US-0227009.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	01-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	05-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	06-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	08-SEP-2000;	2000US-0232081.
PR	12-SEP-2000;	2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234224.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249287.  
 PR 17-NOV-2000; 2000US-0249289.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 06-DEC-2000; 2000US-0251719.  
 PR 06-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM,  
 XX  
 DR WPI; 2001-457726/49.  
 XX  
 PT Isolated polypeptide for treating, preventing and prognosing disorders  
 PT related to the endocrine system including endocrine disorders,  
 PT reproductive disorders, and gastrointestinal disorders and also for  
 PT testing and detection e.g. diagnosis -  
 XX  
 PS Disclosure; SEQ ID No 470; 558pp; English.  
 XX  
 CC The invention relates to cDNAs encoding novel human endocrine  
 CC antigens or a fragment having biological activity, a domain, an epitope,  
 CC full length protein, variant, allelic variant or a species homologue of  
 CC the cDNA/antigen. The DNA and polypeptides are useful for preventing,  
 CC treating or ameliorating a medical condition when administered  
 CC (e.g. by gene therapy or antisense-therapy). Identifying mutations in  
 CC the genes coding for the antigens is useful for diagnosing a pathological  
 CC condition or a susceptibility to a pathological condition. The DNAs,  
 CC antigens and antibodies raised against the antigens useful for treating,  
 CC preventing and/or prognosing disorders related to the endocrine system  
 CC or hormone imbalance or reproductive disorders, cancers of endocrine  
 CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal  
 CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the  
 CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples  
 CC of diseases and disorders are given in the specification. The present  
 CC sequence is genomic DNA fragment form a gene encoding an endocrine  
 CC antigen of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 29543 BP; 6367 A; 8023 C; 7634 G; 7519 T; 0 other;  
 Query Match 3.4%; Score 343.8; DB 22; Length 29543;  
 Best Local Similarity 58.3%; Pred. No. 4.9e-56;  
 Matches 878; Conservative 0; Mismatches 482; Indels 145; Gaps 10;  
 QY 7198 GCCGGCGTGTGAGTACGCGCTGTAATCTTGAACATTGGAGGCTGAGCGAGTGAT 7257  
 Db 5325 GCCAGGCACAGTGGCTCAGCGCTTAATCTTGAACCGTTTGGAGGTCAAGGAGAGAT 5384  
 QY 7258 TGGTTAGCGGCTTGAAGCTTGAAGCCAGGCTGGGCAACATGCAAAACCTCA 7317  
 Db 5385 TGGTTAG-----CCAGAGATTGAAGCAGAGCTGGGCAATGCAAGACCTCG 5435  
 QY 7318 TCTCTCAAAAATAATTAATAATTAATGTCGGGTGTGATGCGCTTCTAATAGTCCCATCA 7377  
 Db 5436 TCTCTAATAAATAAATAAATAATTAATGCGGGGATGTGTGATGCTTGTATGCCAGCTTA 5495  
 QY 7378 CTTCAAGAGCTGAGCCCGGAAAGTGC-----AGCTTCAGTGAAGCCG 7419  
 Db 5496 CTGGGAGGCTGAGCGCGAGATGCTTGAAGCTTGAAGGCTGCGGTAGCTG 5555  
 QY 7420 TGAATGCTACTGCACTTCAAGCTTGGGTGACAGATGACCAATGCTCAAAAAACA 7479  
 Db 5556 TGAATGCTCACTACATTCAGACATGGGTGACAGATGACCAATGCTCAAAAAATA 5615

```

Qy 7480 AAAACAAAAACAAAAA-----ACAAAACAAACAAAA 7517
Db 5616 AATGCAAGCTGAAGGCCAGTGCATCTGGCCCTGCATCTGAAATGACATCACAG 5675
Qy 7518 AACCCATATATATATATATATACCTAGCTAGTGAATGACATATTTGGTAAATCA 7577
Db 5676 ACTTTCGACGCTCATCTGAGCCACACAGGCTGGATTAAGGTGAACAAATGAAGCA 5735
Qy 7578 CCAACATGACCCAGCTACAGATGGGGCAGTCCCTCCCTCTCATGCTAAATTTTCTT 7637
Db 5736 CTGGCTTAGGGTAAATCTAAGGGACACACAAATTTAGGGGACACAGCAAA 5795
Qy 7638 TCTCTGACTCAAGTCTTGTGT-----TGTGTCCTGTGTGTGAGA 7681
Db 5796 ATTGATTAATCAAGATTAAGATTTTAAATGCAATGAGTTTGTGTTTTTGTAGA 5855
Qy 7682 TGAAGTCTACTC-TGTACCCAGGCTGAGTGCATAGGGCAATCTTGTTGCTGCAA 7740
Db 5856 TGGAGTTCACCTCTGTGCTGCCAGCTGAGTGCATAGGCTGATCTCAACTCACTGAAA 5915
Qy 7741 CCTCTGCTCTGGGTTCAAGCATCTCTGCTCAGCCCTCCGTTATAGCTGGACTAC 7800
Db 5916 CTTCTGCTCTCCGGGTTCAAGCATCTCTGCTCAGCCCTCCGAGTAGTGGATTTAC 5975
Qy 7801 AGGCGCATACCAACCATCTGGCTATTTTGTATTTTGTGTTTGGTTACAAATGACTAT 7860
Db 5976 AGGTGTGCGCATATGCGCATATTT-----6003
Qy 7861 TTATTAATTAATTTTGTATTTTGTATTTTGTATAGATAGAGTTTACATGTTGGCCAGCTG 7920
Db 6004 -----TTTTGTATTTTGTATAGATAGAGTTTGTCTCATGTTGTCCAGCTG 6051
Qy 7921 GTCTGAATCTCTGACCTCAGGTGATCGGCTGCTGGGCTCCCAAGTCTGAGATTA 7980
Db 6052 GTCTGAATCTCTGACCTCAGGTGATCGGCTGCTGGGCTCCCAAGTCTGAGATTA 6111
Qy 7981 CAGGATGAGCAACACGCTGCGCTGAGGTTT-ATCTATCTCTTCTTCTCT 8038
Db 6112 CAGGATGAGCAACACGCTGCGCTGAGGTTT-ATCTATCTCTTCTCTCT 6171
Qy 8039 TCACACTTGGCTTGACGTCGACATCATGTTCTCCACTTCTCACTACTTCAATGA 8098
Db 6172 AAAAAATCCAAATGAACAAATATCAAAATTTAAATGAATGAGATGACATGAGGG 6231
Qy 8099 TCTTTCAGTCTCAGTCCACTGATACCTCCCTCAGTTGCTTTTCTCAGTAAATTT 8158
Db 6232 AATATGAAAAACATGCAACCTCTTAAACATGTTGTGTCTATTTTAAATGTTGATTT 6291
Qy 8159 TCCAGAGAGGGAATCTGAATGGCCAGTCCATATTT-----TTCAGACACACCAATTA 8212
Db 6292 TTTTTCAGAGGATTAAGATGTTAAATAATTTTAAACATGAGAGATGATGATTA 6351
Qy 8213 AAGTGTGATGTCGACGCTATGATTTGGCTACATTAATGGGTTGGAACTCAT-CA 8268
Db 6352 AAGTGTGATGTCGACGCTATGATTTGGCTATTAACAAACAGAACTTATATATCT 6411
Qy 8269 TTTACTCATTCGACAAAGCAGATAGCTGTCTCAAAATAGGCCCTCTGGGCCAGG 8328
Db 6412 TTTGCTTCAAGTTGAATGATTAAGAAATTAAGTAAAGCCGGGAGGTGGCTTGGCCGGG 6471
Qy 8329 TGTGTGCTCATGCTCTATATATCCCAACACTGTGGGAGGCCGAGGGGGGAGATCACTTG 8388
Db 6472 TGTGTGCTCATGCTCTATATATCCCAACACTGTGGGAGGCCGAGGGGGGAGATCACTTA 6531
Qy 8389 AGTCCAGAGATTTAGACACAGCTGCGGACATGTTGAATCTCATCTTAATAAATA 8448
Db 6532 AGGTCCAGAGATTTAGACACAGCTGCGGACATGTTGAATCTCATCTTAATAAATA 6591
Qy 8449 CAAAAAATTTAGCCAGGTGTGGTGCATGACACAGTATGCCAGTGTTCAGAGAGGCTGAG 8508
Db 6592 CAAAAAATTTAGCC-----AGCCAGAGGTGTGTGGGAGGCTTAC 6628
Qy 8509 GTGGAGGATTTGCTGAGTGTGGAGGACAGATTTGACATGTAACCGTACGTGCTCTG 8568

```

```

Db 6629 ACAGAGAAATCACTTAACACAGGAGGTGAGGTTGACAGTGCAGATCGTCCACAG 6688
Qy 8569 CAATCCAGCTGTGGTGAAGATTTAGACACCTGTCTCAAAAAAACAATTAATTAATAA 8628
Db 6689 CCCTCCAGCTGTGGTGAAGATTTAGACACCTGTCTCAAAAAAACAATTAATTAATAA 6748
Qy 8629 TAAAT 8633
Db 6749 TTATTT 6753

RESULT 13
ABT10147/c
ID ABT10147 standard; cDNA, 185035 BP.
XX
XX ABT10147;
AC
XX
XX 04-DEC-2002 (first entry)
DT
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 281.
XX
XX Human, breast specific gene, breast cancer; differential expression;
XX cytosolic; gene therapy; gene; ss.
XX
XX Homo sapiens.
XX
XX NO200259271-A2.
XX
XX 01-AUG-2002.
XX
XX 25-JAN-2002; 2002MO-US02176.
XX
XX 25-JAN-2001; 2001US-263757P.
XX 25-APR-2001; 2001US-286090P.
XX 23-MAY-2001; 2001US-292517P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Orr MS, Nacion M, Diggins JC, Zeng W;
XX
XX WPI; 2002-674803/72.
XX
XX Diagnosing breast cancer in a patient comprises detecting the level of
XX PT gene expression in cell or tissue samples, where a differential gene
XX expression is indicative of breast cancer.
XX
XX Claim 1; SEQ ID NO 281; 260bp + Sequence Listing; English.
XX
XX The present invention relates to methods of diagnosing breast cancer in a
XX CC patient, which comprise detecting the level of expression in a tissue
XX sample of two or more genes selected from those shown in ABT09867-
XX CC ABT1112, where a differential expression of the genes indicates breast
XX CC progression, and in monitoring treatment of breast cancer in patients.
XX CC The methods are also useful as a screening tool for agents that modulate
XX CC the onset or progression of breast cancer. The breast cancer genes may be
XX CC used as diagnostic markers for the prediction or identification of the
XX CC malignant state of breast tissue, for confirming the type and progression
XX CC of cancer, and for drug screening and assays. The present sequence is a
XX CC coding sequence of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ffp.wipo.int/publ/published_pct_sequences.
XX
XX Sequence 185035 BP; 42256 A; 51727 C; 51210 G; 39842 T; 0 other;
SQ
Query Match 3.4%; Score 343.8; DB 24; Length 185035;
Best Local Similarity 58.3%; Pred. No. 6,9e-56;
Matches 878; Conservative 0; Mismatches 482; Indels 145; Gaps 10;
Qy 7198 GCCGGGCGTGTGAGTCAAGCTGTAATCTAGCACTTTGGAGGCTGAGGCGAGTGGAT 7257

```

D	37219	GCGAGGCA	CAGTGGCT	CACGCGCT	TAATCA	CTCAGCGCTT	TGGAGGTCA	GAGCAGAGAT	37166					
Q	7258	TGCTTGA	CGGCTT	GAGCCT	TAGAGTT	TAGACCA	CGCTGGGCA	CAATGGCAAA	CTCA	7317				
D	37159	TGCTTTGA	AG-----	CCAA	AGTTT	TGAGAC	CAAGCGCT	GGGCACA	TAGCAAGAC	CCCTG	37109			
Q	7318	TCTCTA	CAAAAAT	ATATATAA	AAATTA	TATGTGGG	TGTGTAT	AGTCCG	TCTATAT	GTCCCAT	CTTA	7377		
D	37108	TCTCTA	CTAA	AAAAAT	AAAAA	GAATTT	AGCCGG	AGATGG	TGTGTCA	GTGTAT	GTCCAG	CTTA	37049	
Q	7378	CTTCAG	AGGCT	GAAGCC	CGGAA	AGGTG	CG-----	AGGCTT	CAGT	GAAGCC	G	7419		
D	37048	CTGGG	AGAGG	CTGA	AGGCGG	GAGAT	TGCGTT	TAGAGCT	TGGGAT	TTTGA	AGGCTCG	GTGAC	CTG	36988
Q	7420	TGATCG	TGCTA	CTGCA	CTCCAG	CTGGGT	GA	CAGA	GTGAG	ACCAT	GTCTCA	AAAAA	AACA	7479
D	36988	TGATGT	GGCCACT	CACTTCC	AGCAT	GGGTGA	CAGA	GTGAG	CACTGT	CTCTAA	TAAATA			36922
Q	7480	AAAA	CAAAAA	CAAAAA	CAAA-----	ACAA	CAAA	CAAA	CAAA	7517				
D	36928	AATGCA	AGCTGA	AAAGGCC	AGTGCAT	CTGGGCG	CTGCAT	CTGAAAT	TGACAT	CAACG				36868
Q	7518	AACCAT	ATATAT	ATATAT	ATACCT	AGCTGA	GGTGA	ATGCA	CTATTT	TGTAA	ATA	7577		
D	36868	ACTTT	CTGCA	CCCTCA	TCTCAG	GCAC	AGAGG	CTGGGAT	TAGG	GTGA	AAACAG	TAAGCA		36809
Q	7578	CCAA	CATGAC	CCAGCTA	CAGAT	TGGGG	CAGTCC	TCCCTCA	CTCAT	GTGAA	ATTTT	CTT		7637
D	36808	CTGCTT	AGGGTAT	AAATCT	TAAGGG	AGAC	CAAGAA	ATTTCA	AGGGGAC	ACAGAA				36749
Q	7638	TCTCTG	ACTAC	AGTTT	GTGT-----	TGTT	GTCTGT	GTGT	TGAG					7681
D	36748	ATTGAT	GTATCA	MAATTA	ATGA	TGTTTT	TAATGCA	TGTGT	TTTGT	TTTGT	TTTGT	TGAGA		36889
Q	7682	TGAGT	CTCA	CTC-TG	CACCCAG	GGCTGA	GTGCAAT	GGGCAAT	CTGTG	CTAC	TCAGAA			7740
D	36688	TGAGTT	CTCACT	TGTGTG	CGCCAG	GTGAGT	GCATAG	GGGTGAT	CTCA	CTCA	CTGAAA			36629
Q	7741	CCTCTG	CTCTG	GGTTC	CAAGCA	TCTCTG	CTCAG	CTCCCG	TATAG	TGGGAG	CTAC			7800
D	36628	CCTCG	TCTCCCG	GGTTC	CAAGCA	TCTCTG	CTCAG	CTCCCG	TATAG	TGGGAG	CTAC			36568
Q	7801	AGGCG	CAATCA	CAATG	CTG	GTCTA	ATTTT	TGATTTT	TGGG	TACAT	GTAT			7860
D	36568	AGGTG	TCGCG	CACTAT	CTCCAG	CTAAT-----								36541
Q	7861	TTATTA	TTTAAT	TTTTT	TGTAAT	TATAGAT	ATAGG	TTTCA	CAATG	TGGC	AGGCTG			7920
D	36540	-----	TTTT	TGTAAT	TTTTTA	GAAGAT	GGGTTTT	CTCA	TGTTG	TCAGG	CTG			36493
Q	7921	GTCTG	CAACT	CTG	ACCTCAG	TGATCCG	CTG	CTCG	CTCG	CTCCAA	AGTCTA	AGGATTA		7980
D	36492	GTCTT	GAAC	TCTGAC	CTTA	AGTGA	TCTAC	CTG	CTCTG	CTCCAA	AGTCTA	AGGATTA		36433
Q	7881	CAGCA	TGAGCA	CAACG	CGCTG	CGCTCA	TAAG	TTTTT--	ATCTA	TTCT	CTTTG	CTTCT		8038
D	36432	CAGCA	TGAGCA	CAACG	CGCTG	CGCTCA	TAAT	TTTTT	AAAAA	ATTA	TATGCA			36373
Q	8039	TCA	CAACT	TTTGG	CTTGA	CGTGA	CAATCAT	GTGTT	CTCTC	CACTT	CTCACT	CTCATGA		8098
D	36372	AAAAA	ATCCCA	MAATGA	CAAAAT	ATTCMAA	TTTTTA	ATAA	AGATC	AGATGA	CTGAGG			36313
Q	8099	TCTT	CACTCT	CAGTTCA	CACTGA	TACCT	CCCTCAG	TGCTCT	TTTTT	CTT	AGTATGA	TT		8158
D	36312	AAAT	GAAAAA	CAATG	CAACCT	CTTA	ACATGTT	CGTGT	CAATTTT	TATAT	GGTTGATTT			36253
Q	8159	TC	CA	GAGAGG	GAATCT	GAATGG	CCAGT	CAAT-----	TC	CA	ACCA	CAACCA	CAATTA	8212
D	36252	TTTT	TC	CA	GAGAGATTA	AGTATTA	AAAAAT	ATTTAA	CAATGA	MAAGT	GTATGATTA			36193
Q	8213	AATG	GTGTTG	ATG	CGAC	CTATAT	GTATGG	CTATAT	TATGG	GTGGAA	CTCAT---	CA		8268
D	3													

Chr	Start (Mb)	End (Mb)	Gene	Transcript	Accession	Length (bp)	GC (%)	GC3 (%)	GC4 (%)	GC5 (%)	GC6 (%)	GC7 (%)	GC8 (%)	GC9 (%)	GC10 (%)	GC11 (%)	GC12 (%)	GC13 (%)	GC14 (%)	GC15 (%)	GC16 (%)	GC17 (%)	GC18 (%)	GC19 (%)	GC20 (%)	GC21 (%)	GC22 (%)	GC23 (%)	GC24 (%)	GC25 (%)	GC26 (%)	GC27 (%)	GC28 (%)	GC29 (%)	GC30 (%)	GC31 (%)	GC32 (%)	GC33 (%)	GC34 (%)	GC35 (%)	GC36 (%)	GC37 (%)	GC38 (%)	GC39 (%)	GC40 (%)	GC41 (%)	GC42 (%)	GC43 (%)	GC44 (%)	GC45 (%)	GC46 (%)	GC47 (%)	GC48 (%)	GC49 (%)	GC50 (%)	GC51 (%)	GC52 (%)	GC53 (%)	GC54 (%)	GC55 (%)	GC56 (%)	GC57 (%)	GC58 (%)	GC59 (%)	GC60 (%)	GC61 (%)	GC62 (%)	GC63 (%)	GC64 (%)	GC65 (%)	GC66 (%)	GC67 (%)	GC68 (%)	GC69 (%)	GC70 (%)	GC71 (%)	GC72 (%)	GC73 (%)	GC74 (%)	GC75 (%)	GC76 (%)	GC77 (%)	GC78 (%)	GC79 (%)	GC80 (%)	GC81 (%)	GC82 (%)	GC83 (%)	GC84 (%)	GC85 (%)	GC86 (%)	GC87 (%)	GC88 (%)	GC89 (%)	GC90 (%)	GC91 (%)	GC92 (%)	GC93 (%)	GC94 (%)	GC95 (%)	GC96 (%)	GC97 (%)	GC98 (%)	GC99 (%)	GC100 (%)	GC101 (%)	GC102 (%)	GC103 (%)	GC104 (%)	GC105 (%)	GC106 (%)	GC107 (%)	GC108 (%)	GC109 (%)	GC110 (%)	GC111 (%)	GC112 (%)	GC113 (%)	GC114 (%)	GC115 (%)	GC116 (%)	GC117 (%)	GC118 (%)	GC119 (%)	GC120 (%)	GC121 (%)	GC122 (%)	GC123 (%)	GC124 (%)	GC125 (%)	GC126 (%)	GC127 (%)	GC128 (%)	GC129 (%)	GC130 (%)	GC131 (%)	GC132 (%)	GC133 (%)	GC134 (%)	GC135 (%)	GC136 (%)	GC137 (%)	GC138 (%)	GC139 (%)	GC140 (%)	GC141 (%)	GC142 (%)	GC143 (%)	GC144 (%)	GC145 (%)	GC146 (%)	GC147 (%)	GC148 (%)	GC149 (%)	GC150 (%)	GC151 (%)	GC152 (%)	GC153 (%)	GC154 (%)	GC155 (%)	GC156 (%)	GC157 (%)	GC158 (%)	GC159 (%)	GC160 (%)	GC161 (%)	GC162 (%)	GC163 (%)	GC164 (%)	GC165 (%)	GC166 (%)	GC167 (%)	GC168 (%)	GC169 (%)	GC170 (%)	GC171 (%)	GC172 (%)	GC173 (%)	GC174 (%)	GC175 (%)	GC176 (%)	GC177 (%)	GC178 (%)	GC179 (%)	GC180 (%)	GC181 (%)	GC182 (%)	GC183 (%)	GC184 (%)	GC185 (%)	GC186 (%)	GC187 (%)	GC188 (%)	GC189 (%)	GC190 (%)	GC191 (%)	GC192 (%)	GC193 (%)	GC194 (%)	GC195 (%)	GC196 (%)	GC197 (%)	GC198 (%)	GC199 (%)	GC200 (%)	GC201 (%)	GC202 (%)	GC203 (%)	GC204 (%)	GC205 (%)	GC206 (%)	GC207 (%)	GC208 (%)	GC209 (%)	GC210 (%)	GC211 (%)	GC212 (%)	GC213 (%)	GC214 (%)	GC215 (%)	GC216 (%)	GC217 (%)	GC218 (%)	GC219 (%)	GC220 (%)	GC221 (%)	GC222 (%)	GC223 (%)	GC224 (%)	GC225 (%)	GC226 (%)	GC227 (%)	GC228 (%)	GC229 (%)	GC230 (%)	GC231 (%)	GC232 (%)	GC233 (%)	GC234 (%)	GC235 (%)	GC236 (%)	GC237 (%)	GC238 (%)	GC239 (%)	GC240 (%)	GC241 (%)	GC242 (%)	GC243 (%)	GC244 (%)	GC245 (%)	GC246 (%)	GC247 (%)	GC248 (%)	GC249 (%)	GC250 (%)	GC251 (%)	GC252 (%)	GC253 (%)	GC254 (%)	GC255 (%)	GC256 (%)	GC257 (%)	GC258 (%)	GC259 (%)	GC260 (%)	GC261 (%)	GC262 (%)	GC263 (%)	GC264 (%)	GC265 (%)	GC266 (%)	GC267 (%)	GC268 (%)	GC269 (%)	GC270 (%)	GC271 (%)	GC272 (%)	GC273 (%)	GC274 (%)	GC275 (%)	GC276 (%)	GC277 (%)	GC278 (%)	GC279 (%)	GC280 (%)	GC281 (%)	GC282 (%)	GC283 (%)	GC284 (%)	GC285 (%)	GC286 (%)	GC287 (%)	GC288 (%)	GC289 (%)	GC290 (%)	GC291 (%)	GC292 (%)	GC293 (%)	GC294 (%)	GC295 (%)	GC296 (%)	GC297 (%)	GC298 (%)	GC299 (%)	GC300 (%)	GC301 (%)	GC302 (%)	GC303 (%)	GC304 (%)	GC305 (%)	GC306 (%)	GC307 (%)	GC308 (%)	GC309 (%)	GC310 (%)	GC311 (%)	GC312 (%)	GC313 (%)	GC314 (%)	GC315 (%)	GC316 (%)	GC317 (%)	GC318 (%)	GC319 (%)	GC320 (%)	GC321 (%)	GC322 (%)	GC323 (%)	GC324 (%)	GC325 (%)	GC326 (%)	GC327 (%)	GC328 (%)	GC329 (%)	GC330 (%)	GC331 (%)	GC332 (%)	GC333 (%)	GC334 (%)	GC335 (%)	GC336 (%)	GC337 (%)	GC338 (%)	GC339 (%)	GC340 (%)	GC341 (%)	GC342 (%)	GC343 (%)	GC344 (%)	GC345 (%)	GC346 (%)	GC347 (%)	GC348 (%)	GC349 (%)	GC350 (%)	GC351 (%)	GC352 (%)	GC353 (%)	GC354 (%)	GC355 (%)	GC356 (%)	GC357 (%)	GC358 (%)	GC359 (%)	GC360 (%)	GC361 (%)	GC362 (%)	GC363 (%)	GC364 (%)	GC365 (%)	GC366 (%)	GC367 (%)	GC368 (%)	GC369 (%)	GC370 (%)	GC371 (%)	GC372 (%)	GC373 (%)	GC374 (%)	GC375 (%)	GC376 (%)	GC377 (%)	GC378 (%)	GC379 (%)	GC380 (%)	GC381 (%)	GC382 (%)	GC383 (%)	GC384 (%)	GC385 (%)	GC386 (%)	GC387 (%)	GC388 (%)	GC389 (%)	GC390 (%)	GC391 (%)	GC392 (%)	GC393 (%)	GC394 (%)	GC395 (%)	GC396 (%)	GC397 (%)	GC398 (%)	GC399 (%)	GC400 (%)	GC401 (%)	GC402 (%)	GC403 (%)	GC404 (%)	GC405 (%)	GC406 (%)	GC407 (%)	GC408 (%)	GC409 (%)	GC410 (%)	GC411 (%)	GC412 (%)	GC413 (%)	GC414 (%)	GC415 (%)
-----	------------	----------	------	------------	-----------	-------------	--------	---------	---------	---------	---------	---------	---------	---------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0236868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 03-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 06-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-457726/49.  
 Isolated polypeptide for treating, preventing and prognosing disorders related to the endocrine system including endocrine disorders, reproductive disorders, and gastrointestinal disorders and also for testing and detection e.g. diagnosis -  
 Disclosure; SEQ ID No 471; 558bp; English.  
 PS The invention relates to cDNAs encoding novel human endocrine  
 XX antigens or a fragment having biological activity, a domain, an epitope,  
 CC full length protein, variant, allelic variant or a species homologue of  
 CC the cDNA/antigen. The DNAs and polypeptides are useful for preventing,  
 CC treating or ameliorating a medical condition when administered  
 CC (e.g. by gene therapy or antisense-therapy). Identifying mutations in  
 CC the genes coding for the antigens is useful for diagnosing a pathological  
 CC condition or a susceptibility to a pathological condition. The DNAs,  
 CC antigens and antibodies raised against the antigens useful for treating,  
 CC preventing and/or prognosing disorders related to the endocrine system  
 CC or hormone imbalance or reproductive disorders, cancers of endocrine  
 CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal  
 CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the  
 CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples



CC of diseases and disorders are given in the specification. The present  
CC sequence is genomic DNA fragment form a gene encoding an endocrine  
CC antigen of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 29521 BP; 6369 A; 8019 C; 7628 G; 7505 T; 0 other;

Query Match	3.4%	Score	342.2	DB	22	Length	29521
Best Local Similarity	58.3%	Pred	No. 1e-55				
Matches	877	Conservative	0	Mismatches	483	Indels	145
						Gaps	10

OY	7198	GC	CGGGGGGGTGGTAGTCAAGCCTGAATCTTAAGCACTTTGGGAAGCGTAGGCAAGTGGAT	7257					
Db	5325	GCCAGGCA	CAGGTGGCTCAGCGCTTATACCTCAGCGCTTTGGGAGGTCAAGGCAAGAGAT	5384					
OY	7258	TG	CTTAGACGGGCTTGAAGCTTAGAGATTGAGACCAAGCCTGGGCAACATGCAAAAACCTCA	7317					
Db	5385	TGCTTTGAAG	-----CCAAAGATTGAGACCAAGCCTGGGCAACATGCAAGACCTG	5435					
OY	7318	TCTCTA	CAAAAATATATAAATTAGTGGGGTGTGGTAGTGGCTTCTATAGTCCCATTA	7377					
Db	5436	TCTCTA	CTAAATAATATAAAATTAGCCGGGAGTGGTGTGATGTGTATGTCCACGTA	5495					
OY	7378	CTTCA	GAGAGCTGAGCCCGGAAGTGTG-----AGGCTTCAGTGAGCG	7419					
Db	5496	CTGGGAGAGCTG	AGCCCGGAGATGGCTTGAAGCCTGGGATTTGAAGCTGTGGGTGAGCTG	5555					
OY	7420	TGATGTG	TCTA	CTGCACTCCAGCCTGGGTGACAGATGAGACCATGTCTCAAAAAAACA	7479				
Db	5556	TGATGTG	TCCCACTCATTTCCAGCAGTGGGTGACAGATGAGACACTGTCTTAATAATA	5615					
OY	7480	AAAA	CAAAAAA	CAAAACAA-----ACAA	CAAA	CAAA	CAAA	CAAA	7517
Db	5616	AATG	CAGCTGA	AAAGGCCAGTGC	CATTTGGGCGCTGCATCTGGAAATGCAATCACAG	5675			
OY	7518	AAC	CAATATATATATATATATACCTAGCTGAGTGAAGATGCATATTTTGGTAAATCA	7577					
Db	5676	ACTTTCT	GCACCTCATCTCAGGCGCACAGAGGCTGGATTTAGGGTGAAACAAAGTAAGCA	5735					
OY	7578	CCAA	CATGACCCAGTACAGATGGGGGAGTCCCTCCCTCTCACTGTAATTTTCTT	7637					
Db	5736	CTCG	TTAAGGTTAATAATCTAAGGGGACACCAACAATTTCTAAGGGGACACCAACA	5795					
OY	7638	TCT	CGACTCACAGTTTGTGT-----TGTGTGCTGTGTGTGGA	7681					
Db	5796	ATT	CAGTTATCAAGTAAATATATATGTTTAAATGCAATGTGGTTTGTGTTTGTGGA	5855					
OY	7682	TGA	GTCTCACTC-TGTCA	CCCCAGGCTGAGTGCATGGCGCAATCTTGTGTTCACTGCA	7740				
Db	5856	TGA	GATTACACTCTTGTGCCCCAGGCTGGAAGTCMAATGGCGTGAATCTCACTCACTGAAA	5915					
OY	7741	CCT	GTGCTCTCGGGTTCAAGGATCTCTGTGCTCAGCCTCCGCTATAGCTGGGACATAC	7800					
Db	5916	CTT	CGGTCTCCCGGGTTCAAGGATCTTCTGTGCTCAGCCTCCGAGTACGTGGGATTAAC	5975					
OY	7801	AGG	CGCATACCA	CAATGCTGGCTAATTTTGTATTTTTTTGGGTTACAATGTACTAT	7866				
Db	5976	AGGT	GTGCGCACTATGCCAGCTAAT-----	6003					
OY	7861	TTAT	TATTAATTTTGTATTTTAAAGATATAGGTTTCAACATTTGGCCAGGCTG	7920					
Db	6004	-----	TTTTGTATTTTAAAGATATAGGATGGGATTTCTCATATTTGTCCACAGGCTG	6051					
OY	7921	GTCT	TGAATCTCTGA	CTCAGTGAATCGGCTGCTCGGCTCCCAAAAGTGTAGATTA	7988				
Db	6052	GTCT	TGAATCTCTGA	CTTAAAGTATTCACCTGCTTGCGCTCCCAAAAGTGTAGATTA	6111				
OY	7981	CAG	GATAGCA	CAAGCGCTTGCGCTCTCATATAGTTTT--ATCATTTCTCTTGTCTCT	8038				
Db	6112	CAG	GATAGCA	CAAGCGCGGCTCTGCACATATGTTTTTAAATAAATCAAAATTAATGCA	6171				

Oy	8039	TCAACA	CTTTGGCGCTTGACAGTGGACCATGAGTCTCTCCACATTTCTCACTACATTCATGA	8098
Ds	6172	AAAAAA	ATCCCAATGAACAAAAATATCAAAATTTTAAATTAAGATGAGATGAACTGAGGG	6231
Oy	8099	TCCTTCAG	CTCAAGTTCCAACTGATACCTCCCTCAGTTGCTCTTTTTCCTAGTAGATT	8158
Ds	6232	AAATGGA	AAAAAACAATGCAACCTCTAAACATGTCGTGTGTATTTTAAATGGTTGATT	6291
Oy	8159	TCCAGA	AGGGGAATCTGAATGAGCCCACTCATATTTTCAGACCA-----CACCAATT	8211
Ds	6292	TTTTTCC	AGAGGAATTAAGTAATTAATAAAATATTTTAAACAATGAAGAAGTTGATGATTA	6351
Oy	8212	AAAGTGG	TGATTGCCA---GCCATGTATTTGGCTACATTAATGGTTGGAACTCATCA	8268
Ds	6352	AAATGCT	AAATCATTAATTTCTAAGTGTAAATGCTTATTTTAAACAAACAAATCTATTAATCT	6411
Oy	8269	TTTACCT	CATTGCAAAAGCAGCATAGCTCTGTGTTCTCAAAATATAGGCCCTTGGGCAG	8328
Ds	6412	TTTGCCT	CAGTGAATGTAAATTAAGAAATTAAGTTAAGACCGGGCGAGGTGGCTGGCCGG	6471
Oy	8329	TGTGTGG	CTTCATGCTTAATATCCCAACTGTGGAGGCGCCAGGGGGCGAATCATTTG	8388
Ds	6472	TGTGTGG	CTCATGCTCCGTAAATCCCAAGCACTTTGGAGGGGTGACGAGTAGATCACTTA	6531
Oy	8389	AGTCAGA	AGTTCTTACAGCCAGCCTGGGCAACATGATGTAATTCATCTCTACTTAAATA	8448
Ds	6532	AGGTCAG	AGATTCAGATTCAGACCAAGCTCGGCCCAACATGTGTAAACCCTGTCTCTATTAATA	6591
Oy	8449	CAAAAAA	ATTAGCCAGGTGTGGTGCATGACCAAGTAGTCCCAAGCTGTTACAGAGCTGAG	8508
Ds	6592	CAAAAAA	TTAGCC-----AGCGCAGGGTGTCTCGGAGGGCTAAG	6628
Oy	8509	GTGGAGA	GATTTGCTCCAGTGTGGAGGCAAGATTTGAGTGAACCGTGAACCTGTGCTCTG	8568
Ds	6629	ACAGAGA	AATCTCTTGAACCAAGGAGGTGAGGTGAGTGAAGCGGATCGTGCCACAG	6688
Oy	8569	CAATCCAG	CCCTGGGTGACAGATTGAGACCTGTCTCAAAAACCAATATTAATTAATTA	8628
Ds	6689	CCCTCCAG	CCCTGGGTGACAGAGTAGAGCTCATCTCAAAAATTAATAAATTAACATTA	6748
Oy	8629	TAAAT	8633	
Ds	6749	TTATT	6753	
RESULT 15				
ABA20494/C				
ID	ABA20494	standard; DNA; 16877 BP.		
XX	ABA20494;			
AC				
DE	23-JAN-2002	(first entry)		
XX				
XX		Human nervous system related polynucleotide seq ID NO 12825.		
KM		Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;		
KM		immunopressive; antiinflammatory; anti-HIV; antibacterial; vulnery;		
KM		antiParkinsonian; antischling; antianemic; antiarthritic; cancer;		
KM		antifurmeatic; hepatotropic; cerebroprotective; antiinflammatory;		
KM		antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;		
KM		antiparasitic; cardiac; immune disorder; cardiovascular disorder;		
XX		neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.		
OS		Homo sapiens.		
PN	WO200159063-A2.			
XX				
PD	16-AUG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US01334.			
XX				
PR	31-JAN-2000; 2000US-0179065.			

[illegible][illegible]

PS Disclosure; SEQ ID NO 12825; 1701bp + Sequence Listing; English.

XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or uterine; (b)  
CC immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WFO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 16877 BP; 3563 A; 4508 C; 4516 G; 4290 T; 0 other:

Query Match 3.4%; Score 341.4; DB 22; Length 16877;  
Best local similarity 58.5%; Pred. No. 1,3e-55;  
Matches 876; Conservative 0; Mismatches 516; Indels 105; Gaps 12;

QY 7177 CCTATATTAATAAATCCAGAGCCGGGCGTGTGAGTCAAGCTGTATTCAGCACTTT 7236  
DB 10860 CGTAATGTGAAAAGATTTGGCCGGGCGTGTGAGTCAAGCTGTATTCAGCACTTT 10801  
QY 7237 GGGAGGCTAGAGGAGGAGGAGTTCCTGAGCGGCTGAGAGTTGAGAGCAAGCCT 7296  
DB 10800 GGGATTCGAGAGGAGGAGGAGTTCCTGAGCGGCTGAGAGTTGAGAGCAAGCCT 10752  
QY 7297 GGGCAATGAGCAAAACCTCATCTCAAAAATAA-TAAATTTAGTGGGTGTGTAG 7355  
DB 10751 GGTATATAGGTGAAAGCCCATCTCACTAATAAATCAAAAATTTAGCCGGGTGTGTG 10692  
QY 7356 TGGCTTCTATAGTCCCATCTTCTTCAAGAGGCTGAGCCGGAAG-----GTGAG 7405  
DB 10691 CGGGTCTGTATGCTCCAGCTACTCGGAGGCTGAGGCAAGATGGGTGAAACCCGGAGG 10632  
QY 7406 GCTTCAAGAGCCGATGCTGTCTGCTCAAGCTGAGGAGAGAGAGTGAAGCAGT 7465  
DB 10631 TGAATGTAAGCCCAATCAAGCACTGCACTCAGCTGAGTGAAGAGAGTCCCA 10572  
QY 7466 TCTCAAAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACCA 7525  
DB 10571 TCTCAAAAAA--AAAAAAGAAAAAGAAAGCAATTAATGCACTTAATCTAGAACATC 10514  
QY 7526 TATATATATATATATCTGAGTGAAGTGAAGTCACTATTTGGTAAATCACCAATG 7585  
DB 10513 ATCACTTAGCTAGCTACTTAATATGTGCTCAGAAACATTTATTAAGCTTAC-----G 10460  
QY 7586 ACCCACTACAGATGAGGAGAGTCCCTCCCTCTCACTGTAAATTTTCTCTGAC 7645  
DB 10459 GCTGGGCAAAATATATGAGCAAAAGCTATTTTCAATTAATTAAGAAAAAGTTGAAT 10400  
QY 7646 TCACAGTTTGTGTTGT 7705  
DB 10399 CCATGTATTTT-----TTTTTGAAGAGAGTCTCACTGTGTGTGTGTGTGTGTGT 10350  
QY 7706 CTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 7765  
DB 10349 CTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 10291  
QY 7766 CTTCCGCTCAAGCTCCCTGATAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 7825  
DB 10290 TCTCCGCTCAAGCTCCCTGATAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 10231  
QY 7826 ATTTGTATTTTGT 7885  
DB 10231 ATTTGTATTTTGT 7885

DB 10230 AATTTTTTTTTT-----TTTTGTATTTTT 10206  
QY 7886 AGTAGAGATAGGTTTCAACATGTTGGCCAGGCTGTGCACTCTCAACCTCAAGTGA 7945  
DB 10205 AGTAGAGATAGGTTTCAACATGTTGGCCAGGCTGTGCACTCTCAACCTCAAGTGA 10146  
QY 7946 TCCGCTGCTGAGCTCCCAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 8005  
DB 10145 TCTTCCGCTGAGCTCCCAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 10086  
QY 8006 C-----CTCATAGTATTTTATCTATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 8060  
DB 10085 GATCTCATGATATTTATTTAGTATGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 10026  
QY 8061 GACATCATGTTCTCTCCACTTTCTCACTTCACTTCACTTCACTTCACTTCACTTCACT 8120  
DB 10025 TATGGAGACAGTTCCTCCACTGAATATGATCACTTCTCACCATCATTAAGTTCTTTT 9966  
QY 8121 GATACCTCCCTGAGTGTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 8180  
DB 9965 GGGCGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9906  
QY 8181 CCCAGTCCATATTTTCAACCAACCACTTAAGTGTGATTTGCACTTATG----- 8235  
DB 9905 CTCAAGTTTCAACCTTAAGGAGCAAGTCACTTCTGCAAGAACTGAACCTCCGTGACCC 9846  
QY 8236 --TATTGGTACATTAATGAGTGGAACTCATCTTCTTCTTCTTCTTCTTCTTCTTCTT 8293  
DB 9845 CAATATTCCTTACTTCACTGGAATTAAGCCCTGTCCTTATTTATTAAGCTTACTGA 9786  
QY 8294 AGCTGTGTTTC-----AAATTAAGGAGCCCTGAGGCAAGTGTGAGTGTGAGTGTGAGT 8338  
DB 9785 AGGATGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9726  
QY 8339 CATGCTTATTAATCCCAACCTGTGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8398  
DB 9725 CATGCTTATTAATCCCAACCTGTGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9666  
QY 8399 TTCTAAGCAAGCTGAGGCAACATGTGAATCTCATCTTCTTCTTCTTCTTCTTCTTCTT 8456  
DB 9665 TTCAAGATTAAGCTGAGGCAACATGTGAATCTCATCTTCTTCTTCTTCTTCTTCTTCTT 9606  
QY 8457 TAGCAGAGTGTGAGGCAACATGTGAATCTCATCTTCTTCTTCTTCTTCTTCTTCTTCTT 8516  
DB 9605 CCGAGTGTGAGGCAACATGTGAATCTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 9546  
QY 8517 ATTGCTGAGTGTGAGGCAACATGTGAATCTCATCTTCTTCTTCTTCTTCTTCTTCTTCTT 8576  
DB 9545 ATGACTTGAACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9486  
QY 8577 CTTGGGTGAAGATTGAACCTCTGTTCAAAAAACAATTAATTAATTAATTAATTAATTAAT 8633  
DB 9485 CTTGGGTGAAGATTGAACCTCTGTTCAAAAAACAATTAATTAATTAATTAATTAATTAAT 9429

Search completed: December 15, 2003, 12:37:36  
Job time : 1612.46 secs





Db 10845 GGAAGCTGTTGGTGGTTTGGCCAAAGCTTCTTGAAGACTGCACTACGACTGAAATGTCC 10786  
Qy 8064 CATGATGTTCTCTCCACTTCTTCTACTACTGATGATTTTCAAGTCTCAAGTTCAGAT 8123  
Db 10785 CATACGCT 10727  
Qy 8124 ACCTCCCTCAGTGTCT 8183  
Db 10726 ATGGCTCTCCCAAGCTCTCTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 10667  
Qy 8184 AGTCATATTTTTCAGACCAACCACTTAAAGTGTGATTTGCAAGCTTATTTGGCT 8243  
Db 10666 ACACACCTCTTTCAGACAGTATCCCATCTTGTGTCTCTCTCTCTCTCTCTCTCTCTCA 10607  
Qy 8244 ACATATATGGTGGGAACTCATCTTACTTACTTATGCAAAAAGAGAGATAGCTGTGTT 8303  
Db 10606 CACAAGTGAAGAGAGAGAGAGTGA---TACATGGAAGTAAATTTGATGTGATGGTA 10550  
Qy 8304 CTCAAATATAGGGCCCTGGGCGAGTGTGGTCTCATGCTTAAATCCCAAGCTGTGG 8363  
Db 10549 CAAGTAAAAATGAAGTGGCTGGGCGAGTGTGGTCTCAAGCTTGTATCCAGACCTTTGG 10490  
Qy 8364 GAGGCGAGGGGGGAGATCACTTGAATCCAGAGTTTGAACCAAGCTGGGCAATGG 8423  
Db 10489 GAGGCGAGGTGGGTGGATCACTGAGGTGAGAGTTTGAAGAGCTGGCAATATGG 10430  
Qy 8424 TGAATCT 8483  
Db 10429 TGAATCT 10370  
Qy 8484 AGGCCAGCTGTTCAGAGGCTGAGGGGAGTGTGCTGAGTGTGGGAGAGAGAT 8543  
Db 10369 ATGCCAGCTCTTGGAGGCTGAGGAGAGAGATGTCTGAACCCAGAGGAGAGGTT 10310  
Qy 8544 GCAGTAAACCGTACTGTGCTCTGCAATCAGCTGGGTGAGAGATTTGACCTGTCT 8603  
Db 10309 GTATGAGCTAATATGCACTGCACTGCACTGAGCTGGGAGAGAGAGATCTCATCT 10250  
Qy 8604 CAAAAACCAATTAATTAATAA 8626  
Db 10249 CAAAAACCAATTAATTAATAA 10227

RESULT 2  
US-08-965-048-5  
; Sequence 5, Application US/08965048  
; Patent No. 6323244  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Freimer, Nelson  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND  
; FILE REFERENCE: 7853-093  
; CURRENT APPLICATION NUMBER: US/08/965,048  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 45716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-08-965-048-5

Query Match 3.2%; Score 317.2; DB 4; Length 45716;  
Best Local Similarity 62.1%; Pred. No. 1.1e-64;  
Matches 614; Conservative 0; Mismatches 318; Indels 56; Gaps 5;

Qy 7652 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 7711  
Db 20439 TTTATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 20498  
Qy 7712 TGCATGCGCAATCTTGTGTTCACTGCACTGCTCTGAGGTTCAAGCAATCTCTCT 7771

Db 20499 TGCAGTATGCAATCTTGGCTCACTGCAACTCCGCTCTGAGGTTCAAAAGTCTCTCT 20558  
Qy 7772 GCTCAGCTCTCCGTTATAGTGGAGTACAGAGGATACCAACATGCTGGCTAATTTT 7831  
Db 20559 GCTTCAGCTCTCCGTTATAGTGGAGTACAGAGGATACCAACATGCTGGCTCTCTCTCT 20610  
Qy 7832 GTATTTTGTGGTTAAGTATGTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 7891  
Db 20611 -----TAACTTTGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 20632  
Qy 7892 GATAGGTTTTCAGATGTTGGCCAGGCTGGTCTCGAATCTCTGACCTCAGGTGATCCGCC 7951  
Db 20633 GATGGGTTTTCAGAGTTGGCCAGGCTGGTCTCGAATCTCTGACCTCAGGTGATCCGCC 20692  
Qy 7952 TGGCTGGCTCTCCAAAGTCTGAGTTTACAGGCAAGCAACCAAGCTGGCC-CTCA 8010  
Db 20693 TTCTCAGCTCTCCAAAGTCTGAGTTTACAGGTTGAGCCACCAAGCTCTGAGG 20752  
Qy 8011 TAGGTTTATCTATCTCTTGTCTCTCTCACTCTTGGCTGCAAGTGAACATCATG 8070  
Db 20753 TGTGTTTATTTATTTAAGTAAACATTTTATTTCACTTAAATTTCAAGCTTGAATTTA 20812  
Qy 8071 TTCTCTCACTTCTCTCACTTATGATCTTCACTCTCACTCTCACTCTCACTCTCTCTCT 8130  
Db 20813 GATCATGATATTTGAGGTAACCACTCATTTCTGTAAGACTGATGATCTTAATGACTC 20872  
Qy 8131 TCAGTGTCTTTTCT 8190  
Db 20873 AAGACTGAATTTAGCTGAATTAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20932  
Qy 8191 ATTTTCAGACCAACCACTTAAAGTGTGATGTCAGCTCTATGATTTGGCTCTCACTTAA 8250  
Db 20933 CTTAAGACTATCT 20992  
Qy 8251 TGGGTTGGAACTCATCTT-----TACTTATGCAAGAGCAAGATAGCTCTGTTTC 8304  
Db 20993 TAGAGATTAAGTACTTAATGAGCAATCTCTTATGAGAAATTAATGATGATGTGTAG 21052  
Qy 8305 TCATAATAGGCCCC-GGGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8363  
Db 21053 AGATGTAAAGAGGTCAGGCGGCGGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 21112  
Qy 8364 GAGGCGAGGGGGGAGATCATCTGAGTCCAGAGTTTTCAGACCAAGCTGGGCAATGG 8423  
Db 21113 GAGGCGAGGGGAGATCAT--GAGGTCAAGAGTTCAGACCAAGCTGGGCAATGG 21170  
Qy 8424 TGAATCTCATCTCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8483  
Db 21171 TGAATCT 21230  
Qy 8484 AGTCCAGCTGTTCAGAGAGGCTGAGGTGGAGAGATTTGCTGAGTGTGGAGAGGAGAGAT 8543  
Db 21231 AGTCCAGTGTTCAGAGAGGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21290  
Qy 8544 GCAGTAAACCGTACTGTCT 8603  
Db 21291 GCAGTAAACCGATGCACTGCACTCTCAAGCTCTGGGCAACAGTGAAGTCTGTCT 21350  
Qy 8604 CAAAAACCAATTAATTAATAAATAA 8631  
Db 21351 CAAAAACCAATTAATTAATAAATAA 21378

RESULT 3  
US-08-965-048-6  
; Sequence 6, Application US/08965048  
; Patent No. 6323244  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Freimer, Nelson  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND  
; FILE REFERENCE: 7853-093





Oy	7633	TTCTTCTCAGCAACACAGTTTTGTTGGTGTGTGTGTCGTGTGTTGAATGAGACTCAC		7692
Db	53673	ATAATTTAATTTGMACTACTTTTTCTTT-----TTTAGATGAGTCTCAC		53625
Oy	7693	TCTGTCCACGAGCTGGAGTGACAATGGCGAATCTTGGTTCACTGCAACTCTGCCTCT		7752
Db	53624	TCTATCCACCAGGCTGGAGTGACAGTGGACGATCTTGGCTCACTGCAACTCCACTACT		53585
Oy	7753	GGGTTCAAAGGATCCCTCCTGSCCTCACGCCCTCATAGCTGGAGACTACAGGGCATACCA		7812
Db	53564	GGGTTCAAAGAATTTCTCTGCTCCAGCTCTCCTAGTAGCTGGGATTACAGGCAATGTCA		53505
Oy	7813	CCATGCTGGCTAATTTTGTATTTTTTTTTGGGTTAACATGTACTAATTAAATTTAA		7872
Db	53504	CCATGGCCACACTCTTTTGTATTTAAGTAGAG-----		53472
Oy	7873	TTTTTGTATTTTAAGTAGATNAGGTTTCAACATGTGGCCAGGCTGTCTGCACCTCC		7932
Db	53471	-----ACGGGGTYTTTCGCATATGTGGCCAGGCTGGCTCCAATCTCC		53431
Oy	7933	TGACCTCAGGTGATCCGCTCGCTCGGCCCTCCCAAGTCCTAGAGTTACAGGCATAGAGA		7992
Db	53330	TGACCTCAGGTGATCTACACATTTCCGCTCTCCCAAGTGTACAGATTACAGGGTGAAGC		53371
Oy	7993	ACCAGCCTGGCCCCCTCATNAGTTTTTATCTATTTCTCTTGGCTTCTTCAAACTTGGCT		8052
Db	53370	ACCATGCTGGCCGAGACTACTATTT-----TTAATAGGTTTTAATTTCAAAAATAACA		53317
Oy	8053	TGCACGTGAGCAATCATGTTCTCTCCACTTCTTCACTACTTACATTATCTTTCAAGTCTAG		8112
Db	53316	TTTGTCTTTCCAGTGAAGTTTTAGTAAATAGCACCCCAGATCAACCAACATTTGACGA		53257
Oy	8113	TTCCAACTGATACCTCCCTCAGTTGCTCTTTTTTCCTAGTAAGATTTCCAGAGAGGAAT		8172
Db	53256	GAACTCTCAAAATTATTTAACAAGATACTGTGAACAATGGAAAAACAAAACACAAAC		53197
Oy	8173	CTGAATGGCCCACTGCATATTTTTCAGACCAACACACATTAAAGTGGTTGATTGCCAGCT		8232
Db	53196	CAAAAGGAATCATGTTTTCATCCCACATGGGATATGATCCCCCTTGCAAAATCTGACAGAT		53137
Oy	8233	ATGATATTGCTACTATTAAATGGGTGGGAACCTCATATTACTTCAATGGCAAAAGCAGA		8292
Db	53136	TAAGTCTTACTAATTTCTTCCCTGTGTGT-----TTTAACTTGCAAAATCACAGATGT		53082
Oy	8293	TAGCTCTGTTTCTCAAAATAGGGCCCCCTGGCCAGGCTGTGTGCTCATGCTATPATCC		8352
Db	53081	TCATCTTTCTTGCTTAAAACTTTAATACGGCTGGGCGGGTATCTCAGCCTGTAAATCC		53022
Oy	8353	CAAACTGTGGAGGCCGAGGGGGGCGAGATCACTTAATGTCCAGAGATTCTAACCAGCT		8412
Db	53021	CAGACTTTTGGGAGGCCGAATGGGGGATCAC--AAGGTACAGGAATCAAGACATCCT		52964
Oy	8413	GGGCAACATGSTGAATCTCATCTCACTPAATAAATPCAAAATAATTAGCAAGGTGTGTGG		8472
Db	52963	GGCCAACTATGTAAACCCCATCTCTAATAAATAC-AATAATTAGCTGGGGTGTGGTGG		52905
Oy	8473	CATGCAACAGTAGTCCAGCTGTTCAGAGGCTGAGGTGGAGAGATTGTCTCGATGTGGG		8532
Db	52904	TGCATGCTGTGAATCCAGCTAATTCAGAGATGTGAAGGACAAAGAAATTGCTTGAATCAGG		52845
Oy	8533	AGGCACAGATTGCAGTGAACCGTGACTGTGCTCTGTCAATCCAGCTGGGTGAAGATTGG		8592
Db	52844	AGGGGAGGTTTGACGTGAGCCGAGATCGCGCCACTGCACCTCCAGCCT-GGTGACAGAGA		52786
Oy	8593	AGACCCGTGCTCAAAAACAAATTAATAAATTTAAA 8628		
Db	52785	AGATTCGTCTCAAAAACAAACAAAAACCTTTATA 52750		

```

: Patent No. 6323244
:
: GENERAL INFORMATION:
:
: APPLICANT: Chen, Hong
:
: APPLICANT: Fretwiler, Nelson
:
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
: TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
:
: FILE REFERENCE: 7853-093
:
: CURRENT APPLICATION NUMBER: US/08/965,048
:
: CURRENT FILING DATE: 1997-11-05
:
: NUMBER OF SEQ ID NOS: 8
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 5
:
: LENGTH: 45716
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-08-965-048-5

```

Query Match	3.1%;	Score 307.2;	DB 4;	Length 45716;
Best Local Similarity	62.1%;	Pred. No. 2.6e-62;		
Matches 616;	Conservative 0;	Mismatches 318;	Indels 58;	Gaps 6;

[illegible]

QY 2401 ACCGCCACCCGAGACGCTGCTCAGATGCTGAGCTTTGATCTACCACTTGATGAC 2460  
Db 51127 NNN 51068  
QY 2461 AATGGAACCATGCTGTT 2478  
Db 51067 NNN 51050

RESULT 5  
AX345846  
LOCUS AX345846 5234 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 917 from Patent WO0200928.  
ACCESSION AX345846  
VERSION AX345846.1 GI:18493732  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
artificial sequences.

REFERENCE  
1 Olek, A., Piepenbrock, C. and Berlin, K.  
AUTHORS Diagnosis of diseases associated with the immune system  
TITLE Patent: WO 0200928-A 917 03-JAN-2002;  
JOURNAL Epigenomics AG (DE)  
FEATURES  
source 1. 5234  
Location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 1479 a 40 c 1361 g 2354 t  
ORIGIN

Query Match 15.1%; Score 1509.4; DB 6; Length 5234;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 1777; Conservative 0; Mismatches 446; Indels 0; Gaps 0;

QY 12 AGGCAGGCCGAGCAAAATCTGTAGATTCAGACGGGTTCTGACAGCTGAGCAAGTTG 71  
Db 3012 AGGTAGGTTAGTAATTTGTAGATTTGTAGATTTGTAGATTTGTAGATTTGTAGATTTG 3071  
QY 72 TTGAGAAATCTCTGATGAGATCATGGGCTCTCAGAGGAGAAATATAGGTTTCA 131  
Db 3072 TTGAGAAATTTTGTAGATTTGTAGATTTGTAGATTTGTAGATTTGTAGATTTGTAG 3131  
QY 132 AGGCTAGAGGAGAAAGAGTGAAGGAGGAGCTTTAGATATGAGTCCCATTTGCCCA 191  
Db 3132 AGGTTAGAGGAGAAAGAGTGAAGGAGGAGGAGTTTATAGATATGAGTCTTTTATTTAA 3191  
QY 192 CACCCAGAAAGAGACATGCTGCTGCAATGGGAGAAAGTATGATGATTTGGCTGTA 251  
Db 3192 TATTTAGAAAGAGATATGTTTGTATGAGGAGAGAGTATGATATTTGGTTGTA 3251  
QY 252 GCAGCAGTGCATTCGCCAGGCTGCCAAGACTCAGAGATCCAGCCTTGCCCACTGACC 311  
Db 3252 GTAGCAGTGTATTTGTATTTAGGTTTATGAGATTTAGATTTAGTTTGTATTGATT 3311  
QY 312 TATGAGAGGAGATGATTTCAAGCATTTCCTGTAAGTCAAGAGAGAGCATTTGA 371  
Db 3312 TATGAGAGGAGATGATTTATATTTATTTATTTATTTATTTATTTATTTATTTGA 3371  
QY 372 GGCCTGATGCGAGAGGCTGTGATCATTTGTTCCAGAGTTTCCGGAATGTTGTTTCT 431  
Db 3372 GTTTGATGAGAGGTTTGTATTTGTTTATTTAGAGTTTGGAGATGAGTTGTTTCT 3431  
QY 432 GTTGAAGAGAACTTGCAGAGTGAAGAAAGAGCTGAGACTTTTGTGTAATATATAT 491  
Db 3432 GTTGAAGAGAACTTGCAGAGTGAAGAAAGAGATTTGAGATTTTGTGTAATATATAT 3491  
QY 492 GGGAGTGTGAGGGGTCTGAGAGCATCTGTAGAGATTAAGGAGGCTTTCAAGCTTGGTAG 551  
Db 3492 GGGAGTGTGAGGGGTCTGAGAGCATCTGTAGAGATTAAGGAGGCTTTTATTTTATTTAG 3551

QY 552 GGAGCAGGGGTCTGAGACTTCACTGCGCCATAGCTAGTCTGCCATATTTCTTTC 611  
Db 3552 GGAGTACGGGTTTGGAAATTTATTTGTTATATGTTAGTGTGTTATTAATTTTCTT 3611  
QY 612 TGACTCAGTACGCAATCTCAACAGAAATGGGGAGCTTTGGAGTGGGCCAGAGAT 671  
Db 3612 TGAATTAATAGGTAATTTATTAATAGAAATGGGGGTGTTTGGAGTGGGTTTAGAGAG 3671  
QY 672 ACTGAGATAGAGGATGATCCAGAGAGATGAGATGGGGGCGGAGACCTGAGAG 731  
Db 3672 ATTGAGATAGTATGAGATTTTATGAGAGATGAGATGGGGGTGAGATTTAGAGAG 3731  
QY 732 AGAAGAGAGCTGTAGATTAAGAGGCGTGTGACTCTGATCTGATTAATGCTTAC 791  
Db 3732 AGAATAGATTTGTTATGATTAAGGCGGTTGTGATTTTATGANTTTATATGTTATAT 3791  
QY 792 CATACCTTACCCCAATTCCTAATTTCTCTTACCTGAGAGGGGGGAAATGTCAGAA 851  
Db 3792 TATTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 3851  
QY 852 TTTGGCTGAGACATGAGCAACACTACTGACTCTTGAATGCAATTTTGCATTTTTC 911  
Db 3852 TTTGGTTGATATATGATATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3911  
QY 912 TTCAACAATATTTCTGAAACAATCTTATATGCGAGCACTATTTAGAGTCAAGGAT 971  
Db 3912 TTTTAATTAATTTTGTGAAATATTTTATATGTTATGATTTATTTTATTTATTTAT 3971  
QY 972 ATATATGTTAAACAAGAGCAAGCAAAACAAGCAAAACAATCACCAGATTA 1031  
Db 3972 ATATATGTTAAATATATGATGATTAATTAATTAATTAATTAATTAATTAATTA 4031  
QY 1032 GTTACAGATGAAAGAAATTTCAAGTTTGTAGATTAATTAATTAATTAATTAATTA 1091  
Db 4032 GTTACAGATGAAAGAAATTTTATGTTATGATTAATTAATTAATTAATTAATTAAT 4091  
QY 1092 ATGCTAGATTAAGCGCGTCAAGAAAGCTTCAATTTGAAAGTGTAGCAATTTAAGCAGAGT 1151  
Db 4092 ATGCTAGATTAAGCGCGTCAAGAAAGCTTCAATTTGAAAGTGTAGCAATTTAAGCAGAGT 4151  
QY 1152 AGCTAGAAATTTGTGAATTTCCAGTTACAGTTCTATTTGTTGCGGTGTTAAATTA 1211  
Db 4152 AGTTAGAAATTTGTGAATTTTATGTTATGTTTATTTGTTGCGGTGTTAAATTA 4211  
QY 1212 GCTTTTCCCGCAGGAGTAACTACAGAAAGATTAATTTACTAGATGAGGAGTCTC 1271  
Db 4212 GTTTTATTTTATTAAGGTGAATTAATTAAGAAATTAATTAATTAATTAATTAATTA 4271  
QY 1272 TCTGAAGAGAGACACTCTGTTCTGCTCATTACTGTCAACCTTGCATCTCAGAGCA 1331  
Db 4272 TTTGGAAGAGATATTTTGTGTTTGTGTTTATTTATTTATTTATTTATTTATTTA 4331  
QY 1332 CTTTGTCAAGCCCTTTCAGCTGAGAGAGGAGGAGGCTGAGCAATGGGCTTGAC 1391  
Db 4332 TTTTGTGTAAGTTTGTGTTGTTAGTGAAGGAGGAGGAGGTTGAGTATGAGGTTG 4391  
QY 1392 TTTGAAAGAGAGACATTTATTTGCTCCAGACTCTAGCCCAAGGTTAAAGCTAA 1451  
Db 4392 TTTGATTAATGAGAGATTAATTTATTTATTTATTTATTTATTTATTTATTTA 4451  
QY 1452 GGCCTGGGATGCCCGAGAAAGGCGCTGATGAGCTTGAAGAAACCTGCTGAGTA 1511  
Db 4452 GGTGTTGGATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 4511  
QY 1512 TTTCTAAGTATTTATCTGTGTGTGTGTTACTAAAGTATGATTAATTTGCTGTCTA 1571  
Db 4512 TTTTATAGTATTTATTTGTGTGTGTGTGTTATTTAAAGTATGATTAATTTGTTGTTTA 4571  
QY 1572 GCTGCTTGAAGAGGCTTGAACAAGTACAAGCAATTTAGTCCCTCTTCTCAGC 1631  
Db 4572 GTTGTTTTATAGAGGCTTGAATATGATTAATTAATTAATTAATTAATTAATTAAT 4631  
QY 1632 CTCGCCATTTGTGAGATTAATCAATCAAAAGGTGATCTCAGTCTACTCATTCCC 1691

Db 53347 CCATTGCCCAACACCCAGAAAGAGCATGCTCCCTGCATGCGGAGGAGGATGATAGCA 53288  
Qy 241 CATTGGCTGATGAGGAGGATGGCATGGCCAGGCTGCCAAGAGCTCAAGAGTCCAGCTT 300  
Db 53287 CATTGGCTGATGAGGAGGATGGCATGGCCAGGCTGCCAAGAGCTCAAGAGTCCAGCTT 53228  
Qy 301 GCCCACTGACCTATGAGAGGAGGATGATGTTCAACAGACATTTTCAATTCGTAAAGTCAGA 360  
Db 53227 GCCCACTGACCTATGAGAGGAGGATGATGTTCAACAGACATTTTCAATTCGTAAAGTCAGA 53168  
Qy 361 GAGGACATTTGAGCTGATGGCAGAGGCTGTGTGACATGTTGTCAGAGGTTCCGGAATG 420  
Db 53167 GAGGACATTTGAGCTGATGGCAGAGGCTGTGTGACATGTTGTCAGAGGTTCCGGAATG 53108  
Qy 421 TGTGTTTCTGTGTAAGGAACTTCGAGATGAGAAAGGATCGAGACTTTGGTA 480  
Db 53107 TGTGTTTCTGTGTAAGGAACTTCGAGATGAGAAAGGATCGAGACTTTGGTA 53048  
Qy 481 AGATTATATATGAGGATCTGACAGGAGCTGTGAGCCATCTGTGAGGATCAAGGCTTTTCA 540  
Db 53047 AGATTATATATGAGGATCTGACAGGAGCTGTGAGCCATCTGTGAGGATCAAGGCTTTTCA 52988  
Qy 541 GCCTTGGCTAGGAGGAGGAGGCTCTGAACTTCATCTGGCCCATAGCTGAGCTGCCA 600  
Db 52987 GCCTTGGCTAGGAGGAGGAGGCTCTGAACTTCATCTGGCCCATAGCTGAGCTGCCA 52928  
Qy 601 TAATTTCTTTTCTGACCTCACTAGGCAATCTCAACAGAAATGGGAGAGCTTTGGAGTGG 660  
Db 52927 TAATTTCTTTTCTGACCTCACTAGGCAATCTCAACAGAAATGGGAGAGCTTTGGAGTGG 52868  
Qy 661 GCCCAGAGAGTACTGAGATGAGGATGAGATCCAGAGAGAGATGATGAGGAGCCAG 720  
Db 52867 GCCCAGAGAGTACTGAGATGAGGATGAGATCCAGAGAGAGATGATGAGGAGCCAG 52808  
Qy 721 ACACTGAGAGAGAAACAGGACTGTGATGATGAGGAGGCTGTGACTCTGATCTCAT 780  
Db 52807 ACACTGAGAGAGAAACAGGACTGTGATGATGAGGAGGCTGTGACTCTGATCTCAT 52748  
Qy 781 ATGCCATACCTACCTAATACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 840  
Db 52747 ATGCCATACCTACCTAATACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 52688  
Qy 841 ATATGACAGAAATTTGGCTGACACTAGCACTACTAGTACTGATGAAATGCAATTTTG 900  
Db 52687 ATATGACAGAAATTTGGCTGACACTAGCACTACTAGTACTGATGAAATGCAATTTTG 52628  
Qy 901 CATTTTTCATTCACAAATATTTCTGAAACAATCTTATATGCAAGGCACTAATTTAG 960  
Db 52627 CATTTTTCATTCACAAATATTTCTGAAACAATCTTATATGCAAGGCACTAATTTAG 52568  
Qy 961 GAGTCAGGAGATATATATGATGTAACAAGCAGGCAAAACAAAGCAAGCAACAACA 1020  
Db 52567 GAGTCAGGAGATATATATGATGTAACAAGCAGGCAAAACAAAGCAAGCAACAACA 52508  
Qy 1021 TCACAGATATGATGACAGATGAAAGATTTCAAGTTTATGTAAGTAAATATAACAGC 1080  
Db 52507 TCACAGATATGATGACAGATGAAAGATTTCAAGTTTATGTAAGTAAATATAACAGC 52448  
Qy 1081 AAGGGCTGAAATGCTAGATGAGGCGCTCAGAAAGGCTTCATGAGAGAGTACATTT 1140  
Db 52447 AAGGGCTGAAATGCTAGATGAGGCGCTCAGAAAGGCTTCATGAGAGAGTACATTT 52388  
Qy 1141 AAGCAGAGATCAGCTAGAAATATTTGAAATTCAGTTACAGTTATTTGTTCTGGGTT 1200  
Db 52387 AAGCAGAGATCAGCTAGAAATATTTGAAATTCAGTTACAGTTATTTGTTCTGGGTT 52328  
Qy 1201 GGTAAATATAAGCTTTTCCCAAGGTGAAATCTCAAGAAAGCTAATTAATGATG 1260  
Db 52327 GGTAAATATAAGCTTTTCCCAAGGTGAAATCTCAAGAAAGCTAATTAATGATG 52268  
Qy 1261 TGTGTGTCTCTCTGAGAGAGACACTCTGTTTCTGCTCATTAATGTCACCTTTC 1320  
Db 52267 TGTGTGTCTCTCTGAGAGAGACACTCTGTTTCTGCTCATTAATGTCACCTTTC 52208

Qy 1321 ACTTCAGGCACTTTTGGCAAAAGCCCTTTGCCAGTCAAGGAAAGGAGAGGCTGGCATG 1380  
Db 52207 ACTTCAGGCACTTTTGGCAAAAGCCCTTTGCCAGTCAAGGAAAGGAGAGGCTGGCATG 52148  
Qy 1381 GGGCTTGAACATTTGACAAACAGTGAACATTAATTTGCTCCAGACTCACTAGCCCAAGGTT 1440  
Db 52147 GGGCTTGAACATTTGACAAACAGTGAACATTAATTTGCTCCAGACTCACTAGCCCAAGGTT 52088  
Qy 1441 AAAGCTGAAGAGCTTTGGGCAATGCCAAGAGGCCCCGTGATGAAGCTTGAAAGAGCTG 1500  
Db 52087 AAAGCTGAAGAGCTTTGGGCAATGCCAAGAGGCCCCGTGATGAAGCTTGAAAGAGCTG 52028  
Qy 1501 TTCTCTGAGTATTTCTAAGTAAGTTATCTGATGTGTGTTGTTACTTAAAGTAAAGTAT 1560  
Db 52027 TTCTCTGAGTATTTCTAAGTAAGTTATCTGATGTGTGTTGTTACTTAAAGTAAAGTAT 51968  
Qy 1561 TGTGTCTCTAGCTGCTTGAAGCAGAGGCTTGAACACAGTACACAGCAATATTAGTTCCCT 1620  
Db 51967 TGTGTCTCTAGCTGCTTGAAGCAGAGGCTTGAACACAGTACACAGCAATATTAGTTCCCT 51908  
Qy 1621 CCTTTTCTACCTCCCACTTTGTGAGATTAATCAATCAACAAAGGTATCTCAGTCT 1680  
Db 51907 CCTTTTCTACCTCCCACTTTGTGAGATTAATCAATCAACAAAGGTATCTCAGTCT 51848  
Qy 1681 ACTCACTTCCCTGACTTATGATGAGCTGAGACCATTTGCCAGTGTGAGAGTCAAGCTGGA 1740  
Db 51847 ACTCACTTCCCTGACTTATGATGAGCTGAGACCATTTGCCAGTGTGAGAGTCAAGCTGGA 51788  
Qy 1741 CGTCAAGCTGTAGCCCACTTATCTGTTGAAATTCCTGAAGGAGGTTGGGGGCACTG 1800  
Db 51787 CGTCAAGCTGTAGCCCACTTATCTGTTGAAATTCCTGAAGGAGGTTGGGGGCACTG 51728  
Qy 1801 CCGGAAAAAGAGTCTTGATTCAGATTTCTGTCCAGACCTGACCTTATTTGCACTGA 1860  
Db 51727 CCGGAAAAAGAGTCTTGATTCAGATTTCTGTCCAGACCTGACCTTATTTGCACTGA 51668  
Qy 1861 TGTAAATCAGCAATATTTGGCTTATGCTGGGAGACAGACATTTCCAGATGAGTTGAGAG 1920  
Db 51667 TGTAAATCAGCAATATTTGGCTTATGCTGGGAGACAGACATTTCCAGATGAGTTGAGAG 51608  
Qy 1921 TGGGGGTGTGTGTCTGCCAATATATAGGAGGTTCACTGTGACCCAGAGCTGTCC 1980  
Db 51607 TGGGGGTGTGTGTCTGCCAATATATAGGAGGTTCACTGTGACCCAGAGCTGTCC 51548  
Qy 1981 TGTGGCTCTGACAGTCAAGATGAGGCTAGGAGTACTGGAGACACCGTTGACTGGGTTGT 2040  
Db 51547 TGTGGCTCTGACAGTCAAGATGAGGCTAGGAGTACTGGAGACACCGTTGACTGGGTTGT 51488  
Qy 2041 GGAGCTATGCTGTGTCTGAGCAATTTGCCACCCCTTCTCCGTGAGTAAAGCTGGAGC 2100  
Db 51487 GGAGCTATGCTGTGTCTGAGCAATTTGCCACCCCTTCTCCGTGAGTAAAGCTGGAGC 51428  
Qy 2101 TAGAAGCGAAGATGAGTTCTGGGCTAGGAGTAAAGGTCAGTTTATAGGCTCGGT 2160  
Db 51427 TAGAAGCGAAGATGAGTTCTGGGCTAGGAGTAAAGGTCAGTTTATAGGCTCGGT 51368  
Qy 2161 CAAATTTGGGCTCAGGGCTATAGGAAAGGATCGGTCCCAATGATCAAGTATCTATT 2220  
Db 51367 CAAATTTGGGCTCAGGGCTATAGGAAAGGATCGGTCCCAATGATCAAGTATCTATT 51308  
Qy 2221 TTGTTCTCCCTAGAGCTAGTGGCCATGGGAATGTTGGAAGGCAAGCCAGCAAGC 2280  
Db 51307 TTGTTCTCCCTAGAGCTAGTGGCCATGGGAATGTTGGAAGGCAAGCCAGCAAGC 51248  
Qy 2281 CAGAGCTGACTGTAGAGGCTCCTAGACTCTGTCTATCTGTCTGTGTGTCTC 2340  
Db 51247 CAGAGCTGACTGTAGAGGCTCCTAGACTCTGTCTATCTGTCTGTGTGTCTC 51188  
Qy 2341 TGCATTTTATCACTTGTGTTTTTTTTTTTTTTTTTTTTTTTACTTTCATCTCCCT 2400  
Db 51187 TGCATTTTATCACTTGTGTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNNNNNNNNN 51128



[illegible][illegible]

Db 15251 GAGCAAGCAGATGAGGGGGTAACAGCTCTCAACCAATGCTGTGATTGATGGACA 15310  
QY 5297 AAGGTACCGCGAGTGTCTTGGAAGCTTGAGTACGGGAACATGAAGAGAGTCCGGC 5356  
Db 15311 AAGGTACCGCGAGTGTCTTGGAAGCTTGAGTACGGGAACATGAAGAGAGTCCGGC 15370  
QY 5357 CAGCTGTGGGAAGTGTCTCTGCTGCTGAGATGAGTGGGCGCTACTACTGCTTCAGG 5416  
Db 15371 CAGCTGTGGG-ACTGTCTCTGCTGCTGAGATGAGTGGGCGCTACTACTGCTTCAGG 15429  
QY 5417 GTTACCAATCTCTGCTGCTGAGTCAAGGGAGAGTGGTCTCCAGGTACCCGGGG 5476  
Db 15430 GTTACCAATCTCTGCTGCTGAGTCAAGGGAGAGTGGTCTCCAGGTACCCGGGG 15489  
QY 5477 ATGTCCGAGACTCTTCAATGCGCGCTGGCAGAGGTGAGAAAGCCCTGAGACTTGAGA 5536  
Db 15490 ATGTCCGAGACTCTTCAATGCGCGCTGGCAGAGGTGAGAAAGCCCTGAGACTTGAGA 15549  
QY 5537 CCGTGAAGATTGATCCACTTCCCTGAGCTGTGATCTCAAGTGTCTAGCTTCACT 5596  
Db 15550 CCGTGAAGATTGATCCACTTCCCTGAGCTGTGATCTCAAGTGTCTAGCTTCACT 15609  
QY 5597 TTAACTCCGTGTGCGACACCTTGGCCCTTAATCTAGCCCAATTCGATTTTC 5656  
Db 15610 TTAACTCCGTGTGCGACACCTTGGCCCTTAATCTAGCCCAATTCGATTTTC 15669  
QY 5657 CCATTCGCCCTCATATGGGGAAACCAACCCCACTAACCCCAAGCAATCTTCCACTTG 5716  
Db 15670 CCATTCGCCCTCATATGGGGAAACCAACCCCACTAACCCCAAGCAATCTTCCACTTG 15729  
QY 5717 GACCTGACTGACCTCTGCGCTCTTCTGTGTTCTCTGACCCATTTCTCTCCAGG 5776  
Db 15730 GACCTGACTGACCTCTGCGCTCTTCTGTGTTCTCTGACCCATTTCTCTCCAGG 15789  
QY 5777 CATGACACAGGAATGGAGTGGCATGGGAAGATGCCACCAATGGCCCTGATATAG 5836  
Db 15790 CATGACACAGGAATGGAGTGGCATGGGAAGATGCCACCAATGGCCCTGATATAG 15849  
QY 5837 CGGTGAGCCCAATCTAGTCTGTGCTGACATGACGTGACCAATGGTGGCACTAT 5896  
Db 15850 CGGTGAGCCCAATCTAGTCTGTGCTGACATGACGTGACCAATGGTGGCACTAT 15909  
QY 5897 GCGTTGAGTGTGAGAGATGCCCCCAATGCCCAATGTGCTGACATCTCTTTACTT 5956  
Db 15910 GCGTTGAGTGTGAGAGATGCCCCCAATGCCCAATGTGCTGACATCTCTTTACTT 15969  
QY 5957 GTATCTCCCATCTTGAACATTTCTGATGATGATGATGATGATGATGATGATGATG 6016  
Db 15970 GTATCTCCCATCTTGAACATTTCTGATGATGATGATGATGATGATGATGATGATG 16029  
QY 6017 TGTGTCCCATCTTGAACATTTCTGATGATGATGATGATGATGATGATGATGATGATG 6076  
Db 16030 TGTGTCCCATCTTGAACATTTCTGATGATGATGATGATGATGATGATGATGATGATG 16089  
QY 6077 TCTTGTCTGACAGGCGCCAGCCCTTAATCTTAATCTTAATCTTAATCTTAATCTTAAT 6136  
Db 16090 TCTTGTCTGACAGGCGCCAGCCCTTAATCTTAATCTTAATCTTAATCTTAATCTTAAT 16149  
QY 6137 GCTGTGCGCTGTGAGCTCTCTGAGGCGCTTAATGATCTTGTGTGTCTGTGTCTAAT 6196  
Db 16150 GCTGTGCGCTGTGAGCTCTCTGAGGCGCTTAATGATCTTGTGTGTCTGTGTCTAAT 16209  
QY 6197 GCGCTGTGAGCGCTGTGAGCTCTCTGAGGCGCTTAATGATCTTGTGTGTCTGTGTCTAAT 6256  
Db 16210 GCGCTGTGAGCGCTGTGAGCTCTCTGAGGCGCTTAATGATCTTGTGTGTCTGTGTCTAAT 16269  
QY 6257 CTGAGCTCCCTTGTCTTGTGATATCTTGTGATCTCTGAGTCCCGCTGATTTACTT 6316  
Db 16270 CTGAGCTCCCTTGTCTTGTGATATCTTGTGATCTCTGAGTCCCGCTGATTTACTT 16329  
QY 6317 CAGAACTCAATCTTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAG 6376  
Db 16330 CAGAACTCAATCTTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAG 16389

QY 6377 GCCTGGGAGTGAATGTCTCATTTGCTGAGAACACAGCTGAGAGTGTGGTACTTTAGA 6436  
Db 16390 GCCTGGGAGTGAATGTCTCATTTGCTGAGAACACAGCTGAGAGTGTGGTACTTTAGA 16449  
QY 6437 CCTTAGAGGCTGGCTTCACTAGCCCTGTGAGGTTTCTCTCTGAGTAGCCAAATGAGAT 6496  
Db 16450 CCTTAGAGGCTGGCTTCACTAGCCCTGTGAGGTTTCTCTCTGAGTAGCCAAATGAGAT 16509  
QY 6497 ACCCTCCCTTGAACCCGTGGCATCAATTTGGTAAAGCCATCTTAATTAATCTTAGGCTGT 6556  
Db 16510 ACCCTCCCTTGAACCCGTGGCATCAATTTGGTAAAGCCATCTTAATTAATCTTAGGCTGT 16569  
QY 6557 TCTGAGTTCAGTGAAGCAGCAATTAATGATGATGATGATGATGATGATGATGATGATG 6616  
Db 16570 TCTGAGTTCAGTGAAGCAGCAATTAATGATGATGATGATGATGATGATGATGATGATG 16629  
QY 6617 GTGAGCAACCAATCAATCCAACTGAGATATATGATTAATTAAGACAGTGGTAAAGAT 6676  
Db 16630 GTGAGCAACCAATCAATCCAACTGAGATATATGATTAATTAAGACAGTGGTAAAGAT 16689  
QY 6677 ATTAATCGTGAATAATTTTTTTCACAAAAATTTTTTGGCTCTGACCTTGGACAA 6736  
Db 16690 ATTAATCGTGAATAATTTTTTTCACAAAAATTTTTTGGCTCTGACCTTGGACAA 16749  
QY 6737 ATTTGACAGTTATGATCTATCAAGTTCTGTGAAAAATTAATCAACCAATGAGAGCAAA 6796  
Db 16750 ATTTGACAGTTATGATCTATCAAGTTCTGTGAAAAATTAATCAACCAATGAGAGCAAA 16809  
QY 6797 TCTCCACAGAGATTTGACACTATTAATAAGAACATACAGCTAAGTGAACACACACT 6856  
Db 16810 TCTCCACAGAGATTTGACACTATTAATAAGAACATACAGCTAAGTGAACACACACT 16869  
QY 6857 GTTGTAAAAATACAAATTTAACTGAGAACATACAGCTAAGTGAACACATTAAGTCA 6916  
Db 16870 GTTGTAAAAATACAAATTTAACTGAGAACATACAGCTAAGTGAACACATTAAGTCA 16929  
QY 6917 AGAGAACACACAGCCATGTTGGAGGCCATTTGGAGAGACACACAGCAAAAGTGAATGCA 6976  
Db 16930 AGAGAACACACAGCCATGTTGGAGGCCATTTGGAGAGACACACAGCAAAAGTGAATGCA 16989  
QY 6977 GAAAGAGAGAGAGTGAATGAGATGTAATAAGAGGCCACAGAGAACACAGAGAA 7036  
Db 16990 GAAAGAGAGAGAGTGAATGAGATGTAATAAGAGGCCACAGAGAACACAGAGAA 17049  
QY 7037 TAGAGAGACACAGCCATTAAGATCAAGAACTTCAATGAGCCATGAGCCATTAATG 7096  
Db 17050 TAGAGAGACACAGCCATTAAGATCAAGAACTTCAATGAGCCATGAGCCATTAATG 17109  
QY 7097 AGAATGCTAAGTAACTCTTAATGAATAATGATGATGATGATGATGATGATGATGATGATG 7156  
Db 17110 AGAATGCTAAGTAACTCTTAATGAATAATGATGATGATGATGATGATGATGATGATGATG 17169  
QY 7157 CCCACAGATGAGAGAACCTTATTAATTAATAATCCAGGCCGGGCGTGTAGTCAAC 7216  
Db 17170 -CCACAGATGAGAGAACCTTATTAATTAATAATCCAGGCCGGGCGTGTAGTCAAC 17228  
QY 7217 GCGTGTAAATCTAGACCTTTGGAGGCTGAGGAGGTGATGCTTGAAGCGCTTGAAGC 7276  
Db 17229 GCGTGTAAATCTAGACCTTTGGAGGCTGAGGAGGTGATGCTTGAAGC- GCTTGAAGC 17287  
QY 7277 TAGAGTTTGAAGCAGCGCTGGGCAACATGGCAAAACCTCATCTCAACAAAAATTAATA 7336  
Db 17288 TAGAGTTTGAAGCAGCGCTGGGCAACATGGCAAAACCTCATCTCAACAAAAATTAATA 17347  
QY 7337 AATTAGTCGGGTGTGTAGTGGCTTCTTAATGATCCATCTACTTCAAGAGCTGAGCCGG 7396  
Db 17348 AATTAGTCGGGTGTGTAGTGGCTTCTTAATGATCCATCTACTTCAAGAGCTGAGCCGG 17406  
QY 7397 AAGGTGAGGCTTCAATGAGCCGTGATGCTGATCTGACACTCCAGCTGGGTGAGAGAT 7456  
Db 17407 AAGGTGAGGCTTCAATGAGCCGTGAT- GATCTAATGCACTCCAGCTGGGTGAGAGAT 17465



QY	3077	TTCCATACACCTGTATCAAGTCCCTTCATAGGGCAAGGAAATGTGTCCACAAATG	3136
Db	13091	TTCCATACACCTGTATCAAGTCCCTTCATAGGGCAAGGAAATGTGTGTCCACAAATG	13150
QY	3137	GCACCTGGCTAGAAAGACATGTGGAGCATATAATCCAGTCTGTGCTGTCCACCAATGCC	3196
Db	13151	GCACCTGGCTAGAAAGACATGTGGAGCATATAATCCAGTCTGTGCTGTCCACCAATGCC	13210
QY	3197	ATGTGACCCAGGCTGTGTCTGTCTCAGAGGAAGGGGTGCCCTTTTCTTACCTTGCAGAAAGT	3256
Db	13211	ATGTGACCCAGGCTGTGTCTGTCTCAGAGGAAGGGGTGCCCTTTTCTTACCTTGCAGAAAGT	13270
QY	3257	GCCTGTGTGTGGGGAGTCTCGACGTGGCTTTGTTTTCCCTCCGCTCTTTTCTCT	3316
Db	13271	GCCTGTGTGTGGGGAGTCTCGACGTGGCTTTGTTTTCCCTCCGCTCTTTTCTCT	13330
QY	3317	CTCTTCTCAAAATGTCTCATTTCTATCTCAACCAAGTCCCTATGTCTCTTGGGATCATTC	3376
Db	13331	CTCTTCTCAAAATGTCTCATTTCTATCTCAACCAAGTCCCTATGTCTCTTGGGATCATTC	13390
QY	3377	CTAGCCTTTCCATATACCTTCCCTCAGTGATCTCAACCAATCACTTGGCTCTGAGGAATA	3436
Db	13391	CTAGCCTTTCCATATACCTTCCCTCAGTGATCTCAACCAATCACTTGGCTCTGAGGAATA	13450
QY	3437	TCATATGTGTGGACACGTGATCTAGATCTTCTTGAAGTCCAGACATCTCTTCCAT	3496
Db	13451	TCATATGTGTGGACACGTGATCTAGATCTTCTTGAAGTCCAGACATCTCTTCCAT	13510
QY	3497	TGTATGTCTTACAGGACCTTAAATTCAGCATCCCCCAACTTAACTTGTGACTCTCTT	3556
Db	13511	TGTATGTCTTACAGGACCTTAAATTCAGCATCCCCCAACTTAACTTGTGACTCTCTT	13570
QY	3557	ACAAACCAACCTTTCCTCTGTGTGTTCTGTGTTTCAGTAAATGACCCCAAAATGTGCTGA	3616
Db	13571	ACAAACCAACCTTTCCTCTGTGTGTTCTGTGTTTCAGTAAATGACCCCAAAATGTGCTGA	13630
QY	3617	TTTACTAAGAACAAAGTGCACACAGGGGTCTATATCTGGGCTTGGTTATCTTCTAGGT	3676
Db	13631	TTTACTAAGAACAAAGTGCACACAGGGGTCTATATCTGGGCTTGGTTATCTTCTAGGT	13690
QY	3677	TTTATCTCTCCCTGCGCACATTCACCTGTGTGCGCAGCATTCAGATGAGTTTGA	3736
Db	13691	TTTATCTCTCCCTGCGCACATTCACCTGTGTGCGCAGCATTCAGATGAGTTTGA	13750
QY	3737	GCACACTGCTTCTCATGTGTTGGGCTTCGTAGTCTCTCCCTGTGCTGTAACACCCCTT	3796
Db	13751	GCACACTGCTTCTCATGTGTTGGGCTTCGTAGTCTCTCCCTGTGCTGTAACACCCCTT	13810
QY	3797	CTTCACTTGTCAACCTGTGGAAAATTCCTGCTGATTTTTCAAGCTTTTGGGCCCAATGCTTCC	3856
Db	13811	CTTCACTTGTCAACCTGTGGAAAATTCCTGCTGATTTTTCAAGCTTTTGGGCCCAATGCTTCC	13870
QY	3857	TCCTTGGTGTGAACCTTCCACAACCTCTCTAGGACGACTTGAAGGCACTGTGCTATATTC	3916
Db	13871	TCCTTGGTGTGAACCTTCCACAACCTCTCTAGGACGACTTGAAGGCACTGTGCTATATTC	13930
QY	3917	TCAGTGCACTCTTTTACCTTACACTACCTTGTGATGTTGCAATGCTAGAGATTCAGAGATCTT	3976
Db	13931	TCAGTGCACTCTTTTACCTTACACTACCTTGTGATGTTGCAATGCTAGAGATTCAGAGATCTT	13990
QY	3977	CTGTCTTTTGTACAGTGAACCTTCTGGAAGTGAAGAACAAGTCTTGTATTCCTCAAGTCC	4036
Db	13991	CTGTCTTTTGTACAGTGAACCTTCTGGAAGTGAAGAACAAGTCTTGTATTCCTCAAGTCC	14050
QY	4037	CTCACAATGCTGTGGCATATATGTATGTATTCAGAGACGTGTTTCTTGGATGAATGAATGAT	4096
Db	14051	CTCACAATGCTGTGGCATATATGTATGTATTCAGAGACGTGTTTCTTGGATGAATGAATGAT	14110
QY	4097	GAATTAATTAATGAAGAAATGAATGAAGAAATTAACGTATGGGTGATTTGCAGATGAACAG	4156
Db	14111	GAATTAATTAATGAAGAAATGAATGAAGAAATTAACGTATGGGTGATTTGCAGATGAACAG	14170
QY	4157	TTTGCTATATGTTTGTCCACATGATATGTGTGCAATTAATGTGCGCACAGAGTGTCTG	4216

Db	14171	TTGGATATGTTTGTCAACTGATAGTGTGGACATTAATGTGCCAAGAGCTGTCTG	1423
Oy	4217	GGTACAGAGCTAGAGGCAATGTGTATTAGTAATAGTCACTGAAATTTGCCAATCTGAGA	4276
Db	14231	GGTACAGAGCTAGAGGCAATGTGTATTAGTAATAGTCACTGAAATTTGCCAATCTGAGA	1429
Oy	4277	GTGTGTATGTGCAAAAGACACGACATTTGTTGCCACAGATGACCTGAGATGTGTAGG	4336
Db	14291	GTGTGTATGTGCAAAAGACACGACATTTGTTGCCACAGATGACCTGAGATGTGTAGG	1435
Oy	4337	GCCACAGAGGATATCGTATAGCACGTATGATTAATAATGTGTAAATGACAGAGGAC	4396
Db	14351	GCCACAGAGGATATCGTATAGCACGTATGATTAATAATAATGTGTAAATGACAGAGGAC	1441
Oy	4397	AGTATCTGGGAGTGCACAGTCAAAAAGAGTACTTTTGAATGACAGGAGGACAAAGTCTG	4456
Db	14411	AGTATCTGGGAGTGCACAGTCAAAAAGAGTACTTTTGAATGACAGGAGGACAAAGTCTG	1447
Oy	4457	GGTATACCCTCTGTAAGAAGAGGAAAGATACCCAAAGTTGCTCCAGATGATTTCC	4516
Db	14471	GGTATACCCTCTGTAAGAAGAGGAAAGATACCCAAAGTTGCTCCAGATGATTTCC	1453
Oy	4517	TGGAATCCATCCCACTGATGACGTGTGGATATGACCGGTGGAGAATGTAAGCTGA	4576
Db	14531	TGGAATCCATCCCACTGATGACGTGTGGATATGACCGGTGGAGAATGTAAGCTGA	1459
Oy	4577	AGGCGTCCCTCTTCCCAAGGTCAAGTCCAGGCTGSAATCCAGAACCCTGAGATAGTG	4636
Db	14591	AGGCGTCCCTCTTCCCAAGGTCAAGTCCAGGCTGSAATCCAGAACCCTGAGATAGTG	1465
Oy	4637	GGTGTGATGATGCCAGTATGTGTATATGTGATGATGATGTGTGTGTGAGGCCAC	4696
Db	14651	GGTGTGATGATGCCAGTATGTGTATATGTGATGATGATGTGTGTGTGAGGCCAC	1471
Oy	4697	TATGTGCTTTTAAAGAGAGGAAATAGAGACGACGTATGATAGAGTCAAGGTCAAG	4756
Db	14711	TATGTGCTTTTAAAGAGAGGAAATAGAGACGACGTATGATAGAGTCAAGGTCAAG	1477
Oy	4757	GGAAGAGAGGAAGTCAATCTGTGTGAAGGACACTGTGTATTTCTGTGTGAATAGTCCC	4816
Db	14771	GGAAGAGAGGAAGTCAATCTGTGTGAAGGACACTGTGTATTTCTGTGTGAATAGTCCC	1483
Oy	4817	TCATTTGCCCATGACCTTTAGGACAAATCTACCCCTCTTAAATCTTACATACAGTCTC	4876
Db	14831	TCATTTGCCCATGACCTTTAGGACAAATCTACCCCTCTTAAATCTTACATACAGTCTC	1489
Oy	4877	TCCATGEGCCAAATCCCTATTTGGCCCTTCACTTGAATTTATATTAATCTTAACTTAA	4936
Db	14891	TCCATGEGCCAAATCCCTATTTGGCCCTTCACTTGAATTTATATTAATCTTAACTTAA	1495
Oy	4937	ACTAAGCTCCAGAAACCCTATNGCTATTTCTGTACACTCAAGTTGCTCAATGCTTTGGAA	4996
Db	14951	ACTAAGCTCCAGAAACCCTATNGCTATTTCTGTACACTCAAGTTGCTCAATGCTTTGGAA	1501
Oy	4997	TCCTTCTCTCTCGGGGATTCATCTCTCGTTGTGCTCTTTAATTCCTACTTCAATTCAGAT	5056
Db	15011	TCCTTCTCTCTCGGGGATTCATCTCTCGTTGTGCTCTTTAATTCCTACTTCAATTCAGAT	1507
Oy	5057	TCACCTTAAGTATCATTTTCCCTGGGAAGTTTTCACAGACTCTTCCCACTGTGCTTTGCTG	5116
Db	15071	TCACCTTAAGTATCATTTTCCCTGGGAAGTTTTCACAGACTCTTCCCACTGTGCTTTGCTG	1513
Oy	5117	AGCTGATCCGTGTGTGTTTGTCTGTGAATTTTGTGTATGATATCAACCTCCTTTAGGCATC	5176
Db	15131	AGCTGATCCGTGTGTGTTTGTCTGTGAATTTTGTGTATGATATCAACCTCCTTTAGGCATC	1519
Oy	5177	TTCTGTATGTGCTGTAGCTCATGTGTGTGATGATCAATATCTGTGCCCATCTGTGGACCA	5236
Db	15191	TTCTGTATGTGCTGTAGCTCATGTGTGTGATGATCAATATCTGTGTGCATCTGTGGATCCA	1525
Oy	5237	GAGAAAGCATAAAGAGGGCGTATACCGGCTCTACCAAAATGCTGTGATTAATGTGAC	5296



JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 148198)  
Waterston,R.H.  
Direct Submission  
Submitted (28-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jun 14, 2000 this sequence version replaced gi:7715652.

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0223K12  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 142463 bases at least Q40  
Consensus quality: 144172 bases at least Q30  
Consensus quality: 145227 bases at least Q20  
Insert size: 15100; agarose-fp  
Insert size: 147398; sum-of-contigs  
Quality coverage: 4.53 in Q20 bases; sum-of-contigs  
Quality coverage: 4.67 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2755: contig of 2755 bp in length  
2756 2855: gap of unknown length  
2856 12253: contig of 9398 bp in length  
12254 12353: gap of unknown length  
12354 22567: contig of 10214 bp in length  
22568 37602: gap of unknown length  
37603 37702: gap of unknown length  
37703 51053: contig of 13351 bp in length  
51054 51153: gap of unknown length  
51154 66693: contig of 15540 bp in length  
66694 66793: gap of unknown length  
66794 85880: contig of 19087 bp in length  
85881 85980: gap of unknown length  
85981 115955: contig of 29975 bp in length  
115956 148198: contig of 32143 bp in length.

FEATURES  
source

1. 148198  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-223K12"  
1. 2755  
/note="assembly\_name:Contig14"  
2856. 12253  
/note="assembly\_name:Contig15"  
12354. 22567  
/note="assembly\_name:Contig16"  
22668. 37602  
/note="assembly\_name:Contig17"  
37703. 51053  
/note="assembly\_name:Contig18"  
51154. 66693  
/note="assembly\_name:Contig19"

clone\_end:T7  
vector\_side:right  
66794..85880  
/note="assembly\_name:Contig10"  
85981..115955  
/note="assembly\_name:Contig11  
clone\_end:SP6  
vector\_side:right  
116056..148198  
/note="assembly\_name:Contig12"  
BASE COUNT 38577 a 34635 c 33776 g 40404 t 806 others  
ORIGIN

Query Match 75.6%; Score 7556.4; DB 2; Length 148198;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 7636; Conservative 0; Mismatches 121; Indels 7; Gaps 6;

2237 TAGTCCCATGGAATGTTGCTGAAGCGAGCAACCAAGCCAGACGTCGTCGA 2296  
12251 TGGNN 12310  
2297 GGGCCCTGACTCCCTAAGCTGCTTATCTGCTGTTGCTCTGTCATTATACCTT 2356  
12311 NNN 12370  
2357 CTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTTGGCATCTCCCTACCCCAAGAC 2416  
12371 CTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTTGGCATCTCCCTACCCCAAGAC 12430  
2417 GCTGCTCAGATGCTGAGAGCTTTGATGCTACCACTCGATGACAAATGGAACATGCTGT 2476  
12431 GCTGCTCAGATGCTGAGAGCTTTGATGCTACCACTCGATGACAAATGGAACATGCTGT 12490  
2477 TTTTAAAGGTAGAGAGGAGCTGAGGCTTTAGGAGCTTTAGACCTTAACTTCTCTTC 2536  
12491 TTTTAAAGGTAGAGAGGAGCTGAGGCTTTAGGAGCTTTAGACCTTAACTTCTCTTC 12550  
2537 ACAAGGGGTGCTCCTGCTGCTGAGAGGCTTTAGGAATTAATGATGATCACTGACAGC 2596  
12551 ACAAGGGGTGCTCCTGCTGCTGAGAGGCTTTAGGAATTAATGATGATCACTGACAGC 12610  
2597 TTCTCTCAAGCTATCTCAGTAGGCTCAAGGTTTCTCACTGAGGCCCTCAGTAGGTTGGG 2656  
12611 TTCTCTCAAGCTATCTCAGTAGGCTCAAGGTTTCTCACTGAGGCCCTCAGTAGGTTGGG 12670  
2657 TTTTTCAGGGAGCTTTGCTGGAAGAGTCACAAATGGGACCGGAGTTAATCTCAGAGA 2716  
12671 TTTTTCAGGGAGCTTTGCTGGAAGAGTCACAAATGGGACCGGAGTTAATCTCAGAGA 12730  
2717 GATGGAAGATTTCCCAAGCCCTGCTGAGATGCTGCAATTCGTCGAAGTCAACAGTGTCT 2776  
12731 GATGGAAGATTTCCCAAGCCCTGCTGAGATGCTGCAATTCGTCGAAGTCAACAGTGTCT 12790  
2777 TTCTGATCAAGTACTGCTGAGGCAAAATCAGGAGCCAGGCTGGAAGAGGCTGGAATGAC 2836  
12791 TTCTGATCAAGTACTGCTGAGGCAAAATCAGGAGCCAGGCTGGAAGAGGCTGGAATGAC 12850  
2837 ACTGGGAGCCCTTCCCAATAGGCTTGGCATGAGGCCATATAGCATATAGTACAGATT 2896  
12851 ACTGGGAGCCCTTCCCAATAGGCTTGGCATGAGGCCATATAGCATATAGTACAGATT 12910  
2897 TCTTCCCATGATGCTCCCTTCTCTGTAAGAGCTGGGCTTAAGGAGTGTGATGCGTGT 2956  
12911 TCTTCCCATGATGCTCCCTTCTCTGTAAGAGCTGGGCTTAAGGAGTGTGATGCGTGT 12970  
2957 GGGCTGAGAGTGCACATCAGTGCCTTCTTCACTAGTCTTCTTCACTGATTTACACCG 3016  
12971 GGGCTGAGAGTGCACATCAGTGCCTTCTTCACTAGTCTTCTTCACTGATTTACACCG 13030  
3017 CTCTGCTGTAACCTGACATGCTGAGCAATCTTTTCTTCTGCAATGCTCTCCCTTGGC 3076  
13031 CTCTGCTGTAACCTGACATGCTGAGCAATCTTTTCTTCTGCAATGCTCTCCCTTGGC 13090

Db 23737 CTTTGGCTCTTCAAACTTGGCTGACGTGACATCATCTGCTCTCCACTTTCAC 23678  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8148  
 Db 23677 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23618  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8149  
 Db 23617 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23558  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8209  
 Db 23557 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23498  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8269  
 Db 23497 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23438  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8329  
 Db 23437 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23378  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8389  
 Db 23377 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23318  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8449  
 Db 23317 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23258  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8509  
 Db 23257 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23198  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8569  
 Db 23197 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23138  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8629  
 Db 23137 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23078  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8689  
 Db 23077 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 23018  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8749  
 Db 23017 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22958  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8809  
 Db 22957 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22898  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8869  
 Db 22897 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22838  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8929  
 Db 22837 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22778  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 8989  
 Db 22777 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22718  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 9049  
 Db 22717 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22658  
 QY TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 9109  
 Db 22657 TACTTCATGATCTTTGAGTCTCAGTTCGACATGATACCTCCCTCAGTGTCTTTTTC 22598

QY 9169 AGGAGTGGGCGAGGTACTAGACCTGGTAAAGGTTTCAACCATTAAGGGAGTCATC 9228  
 Db 22597 AGGAGTGGGCGAGGTACTAGACCTGGTAAAGGTTTCAACCATTAAGGGAGTCATC 22538  
 QY 9229 AGAAGTCTTGAAGCAGGCTGTGATATATTTTCACTATTTTAAAGATCACTCTG 9288  
 Db 22537 AGAAGTCTTGAAGCAGGCTGTGATATATTTTCACTATTTTAAAGATCACTCTG 22478  
 QY 9289 ACTTTTTCAGAAACATATATTTTAAAGTACAGCATGTAAGGAATCCAGCTAGCA 9348  
 Db 22477 ACTTTTTCAGAAACATATATTTTAAAGTACAGCATGTAAGGAATCCAGCTAGCA 22418  
 QY 9349 ATCCGTGAGTGTCCAAATTTAGAGTGAAGACCGCTTGAATAGAGATAGCAGCAGA 9408  
 Db 22417 ATCCGTGAGTGTCCAAATTTAGAGTGAAGACCGCTTGAATAGAGATAGCAGCAGA 22358  
 QY 9409 GGTGTGAGGAATCAACATGATATATTTTGAAGTGAAGCTGACAGCATTTAATTAAGC 9468  
 Db 22357 GGTGTGAGGAATCAACATGATATATTTTGAAGTGAAGCTGACAGCATTTAATTAAGC 22298  
 QY 9469 TAAAGTGGCCGGGTGTGTGCTTAAGCTGTATCTTACGACTTTGGAGGCCAAGGC 9528  
 Db 22297 TAAAGTGGCCGGGTGTGTGCTTAAGCTGTATCTTACGACTTTGGAGGCCAAGGC 22238  
 QY 9529 GAGTGTGATCACTGAGGTGAGGATTCGAGACCTGATACCAACATGAGTGAACCTCGT 9588  
 Db 22237 GAGTGTGATCACTGAGGTGAGGATTCGAGACCTGATACCAACATGAGTGAACCTCGT 22178  
 QY 9589 CTCTACTAAATATACAAATTTAGCTGGGAATGAGTGGCAATGCTGTATCTCAGCTAC 9648  
 Db 22177 CTCTACTAAATATACAAATTTAGCTGGGAATGAGTGGCAATGCTGTATCTCAGCTAC 22118  
 QY 9649 TTGGAGGCTGAGGCGAGGAATTCGTTGAACCTGGAGGTGAATTTGAGTACCCGA 9708  
 Db 22117 TTGGAGGCTGAGGCGAGGAATTCGTTGAACCTGGAGGTGAATTTGAGTACCCGA 22058  
 QY 9709 GATTGACATTTGACCTCGAGCTGGGGAACAAGAGTGAACCTCGTCTTAATTAATG 9768  
 Db 22057 GATTGACATTTGACCTCGAGCTGGGGAACAAGAGTGAACCTCGTCTTAATTAATG 21998  
 QY 9769 AATGATGATGATATACAGTCAAGATGAGGAAGGGAAGGCTTCAAGAAATGACTAG 9828  
 Db 21997 AATGATGATGATATACAGTCAAGATGAGGAAGGGAAGGCTTCAAGAAATGACTAG 21938  
 QY 9829 CTTTGTGATCTGAGCACTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9888  
 Db 21937 CTTTGTGATCTGAGCACTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 21878  
 QY 9889 AGGAGT 9948  
 Db 21877 AGGAGT 21818  
 QY 9949 CTATTGGCACCCCATAGCATGTCAAGTGAAGCTGATTTGAGACTTAAC 10000  
 Db 21817 CTATTGGCACCCCATAGCATGTCAAGTGAAGCTGATTTGAGACTTAAC 21766

RESULT 3  
 AC024232  
 LOCUS 148198 bp DNA linear HTG 07-JUL-2000  
 DEFINITION Homo sapiens chromosome 11, clone RP11-223K12, WORKING DRAFT  
 SEQUENCE, 9 unordered pieces.  
 ACCESSION AC024232  
 VERSION AC024232.3 GI:8516169  
 KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 148198)  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone

QY 5869 GAGCTGTGACAAACATGCTGCACTATGCTTCACTGTGTGAGAGATGCCCCCACTCCC 5928  
DB 25897 GAGCTGTGACAAACATGCTGCACTATGCTTCACTGTGTGAGAGATGCCCCCACTCCC 25838  
QY 5929 CCAATGTGCTCTCAATCTCTTTTACTTGTATCTCCATCTTGACACATTTCTCAATG 5988  
DB 25837 CCAATGTGCTCTCAATCTCTTTTACTTGTATCTCCATCTTGACACATTTCTCAATG 25778  
QY 5989 TCATGACGTGTGACATTTTGTGCCCCCTGTGCCCCATCTCTGACATGCCCCCTGTGCA 6048  
DB 25777 TCATGACGTGTGACATTTTGTGCCCCCTGTGCCCCATCTCTGACATGCCCCCTGTGCA 25718  
QY 6049 TCCCTCATCTGTGAGGACATATTTCTCAATCTGTGTGCAAGGCCCAAGCCCTTAATTC 6108  
DB 25717 TCCCTCATCTGTGAGGACATATTTCTCAATCTGTGTGCAAGGCCCAAGCCCTTAATTC 25658  
QY 6109 ATTACTGTGTACATCTATCTCCCATGCTGTGCCCCCTGTGACCTCTGTGAGCCCTT 6168  
DB 25657 ATTACTGTGTACATCTATCTCCCATGCTGTGCCCCCTGTGACCTCTGTGAGCCCTT 25598  
QY 6169 ATGACTCTTGTGTCTCTGTGCAATGCCCCCTGTGAGCCCTGTGCTCTGCTGCTC 6228  
DB 25597 ATGACTCTTGTGTCTCTGTGCAATGCCCCCTGTGAGCCCTGTGCTCTGCTGCTC 25538  
QY 6229 CCTGACCTGTATGTGTCTGTGACCTCTGTGCTCTGCTGCTGTGCTGATCTTCTG 6288  
DB 25537 CCTGACCTGTATGTGTGTCTGTGACCTCTGTGCTCTGCTGCTGTGCTGATCTTCTG 25478  
QY 6289 AGTCTGTGCTCTCCCTGTATTTATCTCAGAACTCCATCTTGTGTGAGGTTCTGTGCTT 6348  
DB 25477 AGTCTGTGCTCTCCCTGTATTTATCTCAGAACTCCATCTTGTGTGAGGTTCTGTGCTT 25418  
QY 6349 CATATGCAAGCCCTGTGAGGATGCACTGCTGTGAGGATGAGATGTTCTCATGCTGAGAA 6408  
DB 25417 CATATGCAAGCCCTGTGAGGATGCACTGCTGTGAGGATGAGATGTTCTCATGCTGAGAA 25358  
QY 6409 CCACTGTGAGAGTGTGGTACTTTAGACCTTTAGAGGCTGCTTCACTAGCTGTGAG 6468  
DB 25357 CCACTGTGAGAGTGTGGTACTTTAGACCTTTAGAGGCTGCTTCACTAGCTGTGAG 25298  
QY 6469 GTTTCCTCTGTGAGTGTGAGGATGCACTGCTGTGAGGATGAGATGTTCTCATGCTGAG 6528  
DB 25297 GTTTCCTCTGTGAGTGTGAGGATGCACTGCTGTGAGGATGAGATGTTCTCATGCTGAG 25238  
QY 6529 AAGCCATCTAATTAATCTGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 6588  
DB 25237 AAGCCATCTAATTAATCTGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 25178  
QY 6589 TGCACATGTGAGATATCCCAAGAGAGTGTGAGCAACATCATCACTCAACCTGAGAT 6648  
DB 25177 TGCACATGTGAGATATCCCAAGAGAGTGTGAGCAACATCATCACTCAACCTGAGAT 25118  
QY 6649 TATGTATTAATTAAGACATGTGTGAGATATTAATTAATGTGTGAGATATTTTCACAAA 6708  
DB 25117 TATGTATTAATTAAGACATGTGTGAGATATTAATTAATGTGTGAGATATTTTCACAAA 25058  
QY 6709 ATTTTGTGCTCTGACCTGTGAGCAAAATTTGACAGTATGATCATGAAATTTCTGTTG 6768  
DB 25057 ATTTTGTGCTCTGACCTGTGAGCAAAATTTGACAGTATGATCATGAAATTTCTGTTG 24998  
QY 6769 AAAAATACATCAACATGTGAGCAAAATCTCCACAGAGATGTGCACTATATATAGA 6828  
DB 24997 AAAAATACATCAACATGTGAGCAAAATCTCCACAGAGATGTGCACTATATATAGA 24938  
QY 6829 ACATACATGTGAGATTAACACACCTGTGTGAGAAATCAACATTTAACTGAGAACAT 6888  
DB 24937 ACATACATGTGAGATTAACACACCTGTGTGAGAAATCAACATTTAACTGAGAACAT 24878  
QY 6889 AGCCATGTGAGAACACATTAATGTGAGAACACACAGCATGTGTGAGAACCATTTG 6948  
DB 24877 AGCCATGTGAGAACACATTAATGTGAGAACACACAGCATGTGTGAGAACCATTTG 24818  
QY 6949 GGAGGACACACAGCAAAAGTGAATGCAAAAGAGAGAGAGTGTGAGATTTGTGA 7008

DB 24817 GGAGGACACACAGCAAAAGTGAATGCAAAAGAGAGAGAGTGTGAGATTTGTGA 24758  
QY 7009 AAAACAGGCGCACAGAAACACACAGAAATGAGAGAGACCAAGCCATCTGAGATCAC 7068  
DB 24757 AAAACAGGCGCACAGAAACACACAGAAATGAGAGAGACCAAGCCATCTGAGATCAC 24698  
QY 7069 AGAATTCATGTGCGCATGTGCGCAATATGAGAAATGCTACTGAACTCTTAATGAAAAATGT 7128  
DB 24697 AGAATTCATGTGCGCATGTGCGCAATATGAGAAATGCTACTGAACTCTTAATGAAAAATGT 24638  
QY 7129 CATGTATGTTCCATGACCTGTGAGAGAGCCCAACAGATGAGAGAAACCTTATATTTAA 7188  
DB 24637 CATGTATGTTCCATGACCTGTGAGAGAGCCCAACAGATGAGAGAAACCTTATATTTAA 24578  
QY 7189 AATACCCAGCGCGGAGTGTGAGTACAGCCCTGTATCTTGAAGACCTTGGAGAGCTGAGG 7248  
DB 24577 AATACCCAGCGCGGAGTGTGAGTACAGCCCTGTATCTTGAAGACCTTGGAGAGCTGAGG 24518  
QY 7249 CAGGTGATTTGTGAGCGGCTTGTGAGCTGTGAGAGTGTGAGACCAAGCTGTGAGCAATGCG 7308  
DB 24517 CAGGTGATTTGTGAGCGGCTTGTGAGCTGTGAGAGTGTGAGACCAAGCTGTGAGCAATGCG 24458  
QY 7309 AAAACCTCATCTCTCAAAAAAATATTAATAATGAGTGTGAGTGTGAGTGTGAGTGTG 7368  
DB 24457 AAAACCTCATCTCTCAAAAAAATATTAATAATGAGTGTGAGTGTGAGTGTGAGTGTG 24398  
QY 7369 TCCCATCTATCTTCAAGAGCTGTGAGCCCGGAAAGTGTGAGAGCTTCAATGAGCCGTGATGTC 7428  
DB 24397 TCCCATCTATCTTCAAGAGCTGTGAGCCCGGAAAGTGTGAGAGCTTCAATGAGCCGTGATGTC 24338  
QY 7429 TACTGTACTCAGCCCTGTGAGGAGTGTGAGAGCATGTCTCAAAAAAACAACAAAAACA 7488  
DB 24337 TACTGTACTCAGCCCTGTGAGGAGTGTGAGAGCATGTCTCAAAAAAACAACAAAAACA 24278  
QY 7489 AACAAAAACAACAAAAACAACAAAAACAACAAAAACAATATATATATATATATATATATAT 7548  
DB 24277 AACAAAAACAACAAAAACAACAAAAACAATATATATATATATATATATATATATATAT 24218  
QY 7549 GGTGAGAAATGCACTATTTGTGAGAAATGCAACCAATGAGCCGCTACAGATGAGGAGT 7608  
DB 24217 GGTGAGAAATGCACTATTTGTGAGAAATGCAACCAATGAGCCGCTACAGATGAGGAGT 24158  
QY 7609 CCTCCCTCTCACTGTGAAATTTTCTTCTCTGACTCAAGTTTGTGTTGTGTTG 7668  
DB 24157 CCTCCCTCTCACTGTGAAATTTTCTTCTCTGACTCAAGTTTGTGTTGTGTTG 24098  
QY 7669 CTGTGTTTGTGAGTGTGAGTCTCACTGTGTCAACCAAGCTGTGAGTGTGAGTGTGAGTGT 7728  
DB 24097 CTGTGTTTGTGAGTGTGAGTCTCACTGTGTCAACCAAGCTGTGAGTGTGAGTGTGAGTGT 24038  
QY 7729 GGTTCATGTCAACTGTGCTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 7788  
DB 24037 GGTTCATGTCAACTGTGCTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 23978  
QY 7789 AGCTGAGATCAACAGGCGCATACCAACATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 7848  
DB 23977 AGCTGAGATCAACAGGCGCATACCAACATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 23918  
QY 7849 ACAATGTACTATTTATTAATTTATTTTGTATTTTGTATGAGATGAGGTTTCAACATG 7908  
DB 23917 ACAATGTACTATTTATTAATTTATTTTGTATTTTGTATGAGATGAGGTTTCAACATG 23858  
QY 7909 TTGGCCAGGCTGTGTGCAACTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 7968  
DB 23857 TTGGCCAGGCTGTGTGCAACTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 23798  
QY 7969 GTGTGAGATTAACAGGATGAGAACCAAGCTGTGAGGCTTCAATGTTTATCTATTTCT 8028  
DB 23797 GTGTGAGATTAACAGGATGAGAACCAAGCTGTGAGGCTTCAATGTTTATCTATTTCT 23738  
QY 8029 GTTGTCTTTCACAACTTTGTGAGCTGTGAGCAATGTTCTCTCACTTTCTCAC 8088

Db 28117 GATCGGCGCTTGTATCTTCTCAGGTTTATCTCTCCCTGCGACATTCAGTGTGCG 28058  
Qy 3709 CAGCAATAGCAATCTATCATAGGTTGGAGCACTGCTTCTCTCATGTTTGGGCTGCGAT 3768  
Db 28057 CAGCAATAGCAATCTATCATAGGTTGGAGCACTGCTTCTCTCATGTTTGGGCTGCGAT 27998  
Qy 3769 GCTGCTCCCTCTCTCTGTTAAACACCTTCTCTCATGTTTGGGCTGCGAT 3828  
Db 27997 GCTGCTCCCTCTCTCTGTTAAACACCTTCTCTCATGTTTGGGCTGCGAT 27938  
Qy 3829 TTTTTCAGCTCTTGGGCGCAATGCTTCTCTGTTGTTGAAACCTTCTCTCAT 3888  
Db 27937 TTTTTCAGCTCTTGGGCGCAATGCTTCTCTGTTGTTGAAACCTTCTCTCAT 27878  
Qy 3889 GCGAGCTTAGGAGCTCTGCTATCTTCTGAGTCACTTCTCTCATGTTTGGGCTGCGAT 3948  
Db 27877 GCGAGCTTAGGAGCTCTGCTATCTTCTGAGTCACTTCTCTCATGTTTGGGCTGCGAT 27818  
Qy 3949 TCGATGGCTAGATTCAGAGAGTCTTCTGCTTCTCTCATGTTTGGGCTGCGAT 4008  
Db 27817 TCGATGGCTAGATTCAGAGAGTCTTCTGCTTCTCTCATGTTTGGGCTGCGAT 27758  
Qy 4009 AGACAGAGCTTGTATCTCTCAGTGTCTCTCATGTTTGGGCTGCGAT 4068  
Db 27757 AGACAGAGCTTGTATCTCTCAGTGTCTCTCATGTTTGGGCTGCGAT 27698  
Qy 4069 TGACTGTTTCTTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 4128  
Db 27697 TGACTGTTTCTTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 27638  
Qy 4129 AACGTAATGAGTATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 4188  
Db 27637 AACGTAATGAGTATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 27578  
Qy 4189 GCGATGAATGTCACAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4248  
Db 27577 GCGATGAATGTCACAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27518  
Qy 4249 ATAGTACTGAGTATGCAAACTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4308  
Db 27517 ATAGTACTGAGTATGCAAACTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27458  
Qy 4309 GTCACAGATGAGTATGCAAACTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4368  
Db 27457 GTCACAGATGAGTATGCAAACTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27398  
Qy 4369 TAAAAAATGTTGTAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4428  
Db 27397 TAAAAAATGTTGTAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27338  
Qy 4429 ACTTTTGAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4488  
Db 27337 ACTTTTGAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27278  
Qy 4489 ACCCAAGTGTGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4548  
Db 27277 ACCCAAGTGTGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27218  
Qy 4549 AATGTACCGTGAAGATGTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4608  
Db 27217 AATGTACCGTGAAGATGTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27158  
Qy 4609 TGAATTCGAAGATGTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4668  
Db 27157 TGAATTCGAAGATGTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27098  
Qy 4669 GTGTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4728  
Db 27097 GTGTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27038  
Qy 4729 GCCACATATGCTTGAAGTCACTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4788  
Db 27037 GCCACATATGCTTGAAGTCACTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26978

Qy 4789 CTGTGATATCTGATGTAATGATGCTTCTCATATGTTTCCCATGACCTTGAACAATCT 4848  
Db 26977 CTGTGATATCTGATGTAATGATGCTTCTCATATGTTTCCCATGACCTTGAACAATCT 26918  
Qy 4849 ACCCTCTTGAATGCTTCAATCAAGTGTCTGATGAGCAATCCCTTATGAGCCCTTCACT 4908  
Db 26917 ACCCTCTTGAATGCTTCAATCAAGTGTCTGATGAGCAATCCCTTATGAGCCCTTCACT 28688  
Qy 4909 TTTTCTTATATTTTATCTTTTATCTTTTATCTTTTATCTTTTATCTTTTATCTTTTATCT 4968  
Db 28687 TTTTCTTATATTTTATCTTTTATCTTTTATCTTTTATCTTTTATCTTTTATCTTTTATCT 28678  
Qy 4969 TACACTGATGTTGCTGATGCTTGTGATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5028  
Db 28677 TACACTGATGTTGCTGATGCTTGTGATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28678  
Qy 5029 GTGTGCTTGTATATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 5088  
Db 28677 GTGTGCTTGTATATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 28678  
Qy 5089 TCCCAAGCTCTCCCACTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5148  
Db 28677 TCCCAAGCTCTCCCACTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28618  
Qy 5149 GGTGTAATGATCACTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5208  
Db 28617 GGTGTAATGATCACTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28558  
Qy 5209 ACCATATATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCG 5268  
Db 28557 ACCATATATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCG 28498  
Qy 5269 CACCAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5328  
Db 28497 CACCAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 28438  
Qy 5329 ACCGGAACATGAGAGAGGCTTCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCT 5388  
Db 28437 ACCGGAACATGAGAGAGGCTTCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCT 28378  
Qy 5389 TGGCTGGGCGCTACTACTGCTTCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCT 5448  
Db 28377 TGGCTGGGCGCTACTACTGCTTCTGCGCATCTGCGCATCTGCGCATCTGCGCATCTGCGCATCT 28318  
Qy 5449 GGAAGGCTGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5508  
Db 28317 GGAAGGCTGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 28258  
Qy 5509 AGAGGTGAGAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5568  
Db 28257 AGAGGTGAGAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 28198  
Qy 5569 GTGAT 5628  
Db 28197 GTGAT 28138  
Qy 5629 TCTAGCCCATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5688  
Db 28137 TCTAGCCCATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 28078  
Qy 5689 ACTTAACCCAGCATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5748  
Db 28077 ACTTAACCCAGCATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 28018  
Qy 5749 TTTCTCTCAACATTTCTCTCTCAGGCGATGATGATGATGATGATGATGATGATGATGATGAT 5808  
Db 28017 TTTCTCTCAACATTTCTCTCTCAGGCGATGATGATGATGATGATGATGATGATGATGATGAT 25958  
Qy 5809 CAGTACCAACATGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5868  
Db 25957 CAGTACCAACATGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25898

QY	1501	TTCTCTAGAGTATTTCTTAAGTAAGTTTATCTGTGTGTGGTTTACTAAAGTAGTAATAT	1560
Db	30277	TTCTCTAGTATTTCTTAAGTAAGTTTATCTGTGTGTGGTTTACTAAAGTAGTAATAT	30211
QY	1561	TGCTGTCTTAGAGCTGCCTTAGAGGAGGAGCTTGACAGAGTACACAGCAATATTAGTCCCT	1620
Db	30217	TGCTGTCTTAGAGCTGCCTTAGAGGAGGAGCTTGACAGAGTACACAGCAATATTAGTCCCT	30155
QY	1621	CTTTTCTCACTTCCCCCATTTGTGAGATTAACCTCAATCACAAAAGGTGATCCTCACTCT	1680
Db	30157	CTTTTCTCACTTCCCCCATTTGTGAGATTAACCTCAATCACAAAAGGTGATCCTCACTCT	30099
QY	1681	ACTCACTTCCCTGACCTTATGGAATCCCTGAGACCCATTCGCACTGTGAAGTCAACAGCTGGA	1740
Db	30097	ACTCACTTCCCTGACCTTATGGAATCCCTGAGACCCATTCGCACTGTGAAGTCAACAGCTGGA	30033
QY	1741	CGTCAGCAGTGTACCCAGATTACGTCTTGAAATTTGCTGAAGAGGGGTTGGGGGCACTG	1800
Db	30037	CGTCAGCAGTGTACCCAGATTACGTCTTGAAATTTGCTGAAGAGGGGTTGGGGGCACTG	29977
QY	1801	CCGGGAAAAGAGACTTTTGATTCAGATTTCTGTCCAGACCCGTGACCTTATTGCACTGA	1860
Db	29977	CCGGGAAAAGAGACTTTTGATTCAGATTTCTGTCCAGACCCGTGACCTTATTGCACTGA	29911
QY	1861	TGTATATAGCCCAATATTGGCTTACTCTCTGGGAGACAGCACTTCCCACTGATAGTTGAGG	1920
Db	29917	TGTATATAGCCCAATATTGGCTTACTCTCTGGGAGACAGCACTTCCCACTGATAGTTGAGG	29855
QY	1921	TGGGGGTGTGTCTCTGCAACTCTATATAGGAGATTCAACTGTGTCAACCCAGACTGTCC	1980
Db	29857	TGGGGGTGTGTCTCTGCAACTCTATATAGGAGATTCAACTGTGTCAACCCAGACTGTCC	29799
QY	1981	TGTGGCCTCTGACAGCTCAGCATGGCTTAGGGTACTGGAGACACCCGTTGCACTGGGGTTGT	2040
Db	29797	TGTGGCCTCTGACAGCTCAGCATGGCTTAGGGTACTGGAGACACCCGTTGCACTGGGGTTGT	29733
QY	2041	GGAGCCTTATGCTGTCTTGTGGCAATTGGCACCCCTCTCTCTCGTAGTAAGCTGGGAC	2100
Db	29737	GGAGCCTTATGCTGTCTTGTGGCAATTGGCACCCCTCTCTCTCGTAGTAAGCTGGGAC	29677
QY	2101	TAGAAGCGAAGATTTAGTTCTGGGCTTAGGGTAAAGTAAAGGCGAGTTTAAAGGCTCGGT	2160
Db	29677	TAGAAGCGAAGATTTAGTTCTGGGCTTAGGGTAAAGTAAAGGCGAGTTTAAAGGCTCGGT	29611
QY	2161	CAAAATTTGGGGTCAGGGGCTATGGGAAAGGGATCGGTCCCAATGATCAAGATTTCTATT	2220
Db	29617	CAAAATTTGGGGTCAGGGGCTATGGGAAAGGGATCGGTCCCAATGATCAAGATTTCTATT	29555
QY	2221	TTGTCTTCTCCTAGAGCTATGTGCCCATATGGGAATGTTGCTGAAGGCGAGACCAAGCCAGAC	2280
Db	29557	TTGTCTTCTCCTAGAGCTATGTGCCCATATGGGAATGTTGCTGAAGGCGAGACCAAGCCAGAC	29499
QY	2281	CAGACGTGACGTGTGAGGGCCCTGATCCCTTAAGTCTGTCTTATCTGTCTGGTTGTGTCTC	2340
Db	29497	CAGACGTGACGTGTGAGGGCCCTGATCCCTTAAGTCTGTCTTATCTGTCTGGTTGTGTCTC	29433
QY	2341	TGCAATTTATCACTTCTGTG-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTT	2388
Db	29437	TGCAATTTATCACTTCTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTT	29377
QY	2389	TGCGATCTCCCTACCTCCACCCAGAACGCTGCTCAGATGGCTGAGCTTGATGTACCC	2448
Db	29377	TGCGATCTCCCTACCTCCACCCAGAACGCTGCTCAGATGGCTGAGCTTGATGTACCC	29311
QY	2449	ACCGTGAATGACATNGAACATGCTGTTTTTAAAGTAAAGGAGGAGCTGAGTTAAGGCG	2508
Db	29317	ACCGTGAATGACATNGAACATGCTGTTTTTAAAGTAAAGGAGGAGCTGAGTTAAGGCG	29255
QY	2509	GTTTAAAGACCTTAGACTTACTCTCCTTACAAAGAGGCTCCCTGTCTGTGGAGAGTCTTA	2568
Db	29257	GTTTAAAGACCTTAGACTTACTCTCCTTACAAAGAGGCTCCCTGTCTGTGGAGAGTCTTA	29199
QY	2569	GGAATTAATGATGTATCACTGACAGCTTCTTCAAGCTATCTCAGTAGGTCAAGAGTT	2628

Dd	29197	GGAATTATCTGATGGTATCATCTACAGCTTCCTCAAGCTATCTCAGTAGGTCAAAGTT	2913
Oy	2629	TCTCACTGGGCCCCCTCAGTAGTGTCGTTTTTTCAGGGAGTTTGTAAGAGTCAAC	2688
Dd	29137	TCTCACTGGGCCCCCTCAGTAGTGTCGTTTTTTCAGGGAGTTTGTAAGAGTCAAC	2907
Oy	2689	AAATGGGACCGGGAGTTAATCTCAGAGAGATGGAGAATTTCCCAGCCCTGTGGATGCT	2748
Dd	29077	AAATGGGACCGGGAGTTAATCTCAGAGAGATGGAGAATTTCCCAGCCCTGTGGAGTCT	2901
Oy	2749	GCATTCCGTCAGAGGTCAACAAGTGCTTCTGATCAAGGATCGCTGGGGCCAATAACAG	2808
Dd	29017	GCATTCCGTCAGAGGTCAACAAGTGCTTCTGATCAAGGATCGCTGGGGCCAATAACAG	2895
Oy	2809	GGCCAGGCTGGAAAAGGCTGGAAATCGACACTGGGGACCCTTCCCCAATAGGCTTGGCA	2868
Dd	28957	GGCCAGGCTGGAAAAGGCTGGAAATCGACACTGGGGACCCTTCCCCAATAGGCTTGGCA	2869
Oy	2869	TGAGCCCATAGCATATGGATAGAGAGATTTCTTCCATGTGGCCTCTTCTCTGTAAAG	2928
Dd	28897	TGAGCCCATAGCATATGGATAGAGAGATTTCTTCCATGTGGCCTCTTCTCTGTAAAG	2883
Oy	2929	CTTGGGCTAAGGGAGATGGATGGATGGATGGGCGCTGGAGGGTGAACAATCAAGTGGCTTT	2968
Dd	28837	CTTGGGCTAAGGGAGATGGATGGATGGATGGGCGCTGGAGGGTGAACAATCAAGTGGCTTT	2877
Oy	2989	CTTCAGTCTTAAGTCTTAAGTCTAACACCGCTCTGTACTCTCACACTGTGGCATCTCT	3048
Dd	28777	CTTCAGTCTTAAGTCTTAAGTCTAACACCGCTCTGTACTCTCACACTGTGGCATCTCT	2871
Oy	3049	TTTTTCTCTGGCAATGCTTCTCCCTTGCTTCATGACCTGTATCAAGTCTCTTCTTAG	3108
Dd	28717	TTTTTCTCTGGCAATGCTTCTCCCTTGCTTCATGACCTGTATCAAGTCTCTTCTTAG	2865
Oy	3109	GGCAAGGCAAGTGTGCCAACAACAATGGGACCGCTGTAAAGAGCATATGGAGCATGAA	3168
Dd	28657	GGCAAGGCAAGTGTGCCAACAACAATGGGACCGCTGTAAAGAGCATATGGAGCATGAA	2859
Oy	3169	ATCCAGTCTGCTGTGCTCAACAAGTCCCATGTGACCCAGGCTGTGTGCTCAGAGAGAG	3228
Dd	28577	ATCCAGTCTGCTGTGCTCAACAAGTCCCATGTGACCCAGGCTGTGTGCTCAGAGAGAG	2853
Oy	3229	GGGTGCTTTTCTTCTTACCTTGCCAAAGGTGTGTGTGGGAAAGTCTGACTGTGGC	3288
Dd	28537	GGGTGCTTTTCTTCTTACCTTGCCAAAGGTGTGTGTGGGAAAGTCTGACTGTGGC	2847
Oy	3289	TTTGTGTTTCCCTCTGCTGCTCTTTTCTCTCTCTCTCAAAATGTCTCATTTCTTCAACA	3348
Dd	28477	TTTGTGTTTCCCTCTGCTGCTCTTTTCTCTCTCTCTCAAAATGTCTCATTTCTTCAACA	2841
Oy	3349	GTTCCCTAATGTTCCTTGGGGATGCATCTGACCTTTCATATPACCTTCCCTCAGTATC	3408
Dd	28417	GTTCCCTAATGTTCCTTGGGGATGCATCTGACCTTTCATATPACCTTCCCTCAGTATC	2835
Oy	3409	TCAAACATCACTTGGCTCTGAGGAAATCTATGCTGTGACACTGTGATCTGATCTACT	3468
Dd	28357	TCAAACATCACTTGGCTCTGAGGAAATCTATGCTGTGACACTGTGATCTGATCTACT	2829
Oy	3469	TTTCTGAGTCCAGACATCTCTTTTTCATTTGTATGTTTTCACAGGCACTTAAATTCAGAT	3528
Dd	28297	TTTCTGAGTCCAGACATCTCTTTTTCATTTGTATGTTTTCACAGGCACTTAAATTCAGAT	2823
Oy	3529	CCCCCAACCTAAGCTTTGACATCTCTTTCACAAACCAACTTTCCTCTGTGTTTCTGTT	3588
Dd	28237	CCCCCAACCTAAGCTTTGACATCTCTTTCACAAACCAACTTTCCTCTGTGTTTCTGTT	2817
Oy	3589	TCAGTAAATGACCCCAAAATGTGCTGTATTACTACAAACCAAGTGAACAAGGCTTCAT	3648
Dd	28177	TCAGTAAATGACCCCAAAATGTGCTGTATTACTACAAACCAAGTGAACAAGGCTTCAT	2811
Oy	3649	GATCTGGGCTTGGTTATCTTCAAGTTATATCTCTCCCTGCAATTCACATGTGTGC	3708



```

repeat_region      /rpt_family="L2"
                    9656..9676
                    complement(9677..10005)
repeat_region      /rpt_family="L2"
                    10006..10186
                    /rpt_family="MSRA"
repeat_region      /rpt_family="MSRA"
                    10187..10496
                    /rpt_family="AluSg"
repeat_region      /rpt_family="MSRA"
                    10497..10671
                    /rpt_family="MSRA"
repeat_region      /rpt_family="MSRA"
                    complement(10672..10981)
repeat_region      /rpt_family="AluY"
                    10982..11052
                    /rpt_family="MSRA"
repeat_region      /rpt_family="MSRA"
                    complement(11053..11484)
repeat_region      /rpt_family="L2"
                    11485..12337
                    /rpt_family="FLAM_C"
repeat_region      /rpt_family="FLAM_C"
                    complement(12507..12954)
                    /rpt_family="L2"
repeat_region      12955..13714
                    /rpt_family="L1MA7"
repeat_region      /rpt_family="L1MA7"
                    complement(14073..14372)
                    /rpt_family="AluSg"
repeat_region      14373..14974
                    /rpt_family="MLT1K"
                    complement(14975..15266)
repeat_region      /rpt_family="AluSx"
                    15267..15301
                    /rpt_family="MLT1K"
repeat_region      15302..15440
                    /rpt_family="L1R50"
                    15441..15817
                    /rpt_family="L1R33A"
repeat_region      /rpt_family="L1R33A"
                    complement(15824..15945)
                    /rpt_family="MIR"
repeat_region      15946..16090
                    /rpt_family="AT-rich"
                    complement(16091..16371)
                    /rpt_family="AluJo"

Query Match      99.7%; Score 9973.2; DB 9; Length 112295;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 9997; Conservative 0; Mismatches 3; Indels 12; Gaps 1;

QY      1 TCCCTCTCCCAAGGCGCCAGCAAAATCTGTAGATTCAACAGAGTTCTGACAGCTG 60
DB      31777 TCCCTCTCCCAAGGCGCCAGCAAAATCTGTAGATTCAACAGAGTTCTGACAGCTG 31718

QY      61 AAGACAAGTTTGTGAGAAATTCCTGATGAGATCAATGGGTGCTCAGAGGAGAAATA 120
DB      31717 AAGACAAGTTTGTGAGAAATTCCTGATGAGATCAATGGGTGCTCAGAGGAGAAATA 31658

QY      121 TAAGTTTCAGAGGCTGAGAGGAAAGAAAGGTGAGGAGAGTCTTAGAATAGTGGCTC 180
DB      31657 TAAGTTTCAGAGGCTGAGAGGAAAGAAAGGTGAGGAGAGTCTTAGAATAGTGGCTC 31598

QY      181 CCATTGCCCAACACCCAGAAAGAGCATGCCCTGCAATGGGAGAGAGGTGAGATAGAGA 240
DB      31597 CCATTGCCCAACACCCAGAAAGAGCATGCCCTGCAATGGGAGAGAGGTGAGATAGAGA 31538

QY      241 CATTGGCTGTAGACGATGCGATTGCCAGGCTGCCAAGACTCAGAGAGTCCAGCCTT 300
DB      31537 CATTGGCTGTAGACGATGCGATTGCCAGGCTGCCAAGACTCAGAGAGTCCAGCCTT 31478

QY      301 GCCCACTGACTATGAGAGGAGAAATGATTCAACAGCAATTTTCAATTGTAAGTACAGA 360
DB      31477 GCCCACTGACTATGAGAGGAGAAATGATTCAACAGCAATTTTCAATTGTAAGTACAGA 31418

QY      361 GAGACATTTGAGCTGATGCGAGAGGCTGTGTGACATGTTGTTCCAGAGTCCGGAATG 420
DB      31417 GAGACATTTGAGCTGATGCGAGAGGCTGTGTGACATGTTGTTCCAGAGTCCGGAATG 31358

```

```

QY      421 TGGTTTCTCTGTGAGAGAACTTCGAGATGAGAAAGGATCTGAGACTTTTGGTA 480
DB      31357 TGGTTTCTCTGTGAGAGAACTTCGAGATGAGAAAGGATCTGAGACTTTTGGTA 31298

QY      481 AGATTATATATGAGACTGTCAAGGGGTCTGAGACCATCTGTGAGAGATCAAGGCGCTTCA 540
DB      31297 AGATTATATATGAGACTGTCAAGGGGTCTGAGACCATCTGTGAGAGATCAAGGCGCTTCA 31238

QY      541 GCGTTGGCTAGAGAGAGAGGGGTCTGTGAACTTCACTCCGGCCCTAGAGCTGAGCCCA 600
DB      31237 GCGTTGGCTAGAGAGAGAGGGGTCTGTGAACTTCACTCCGGCCCTAGAGCTGAGCCCA 31178

QY      601 TAATCTTTTCTGACTCACTAGAGCAAAATCTCAACAGAAATGAGGAGGAGCTTTGAGAGTGG 660
DB      31177 TAATCTTTTCTGACTCACTAGAGCAAAATCTCAACAGAAATGAGGAGGAGCTTTGAGAGTGG 31118

QY      661 GCCCAGAAAGTACTGAGATGAGTCAAGGTGAGATCCAGAAAGATGATGTGGGCGCAG 720
DB      31117 GCCCAGAAAGTACTGAGATGAGTCAAGGTGAGATCCAGAAAGATGATGTGGGCGCAG 31058

QY      721 AACTGAGAGAGAGAAACAGAGCTGTCAATAAAGGCGCTGTGACTCTTAGATCTCAT 780
DB      31057 AACTGAGAGAGAGAAACAGAGCTGTCAATAAAGGCGCTGTGACTCTTAGATCTCAT 30998

QY      781 ATGCCCTACTACATACCTACACCCCAATCTTAATATTTCTCCACCTAGAGGAGGAGAA 840
DB      30997 ATGCCCTACTACATACCTACACCCCAATCTTAATATTTCTCCACCTAGAGGAGGAGAA 30938

QY      841 ATTGTGAGAAATTTGGCTGCAACACTAGCAACACTAGTACTGTAATGATGATTTTG 900
DB      30937 ATTGTGAGAAATTTGGCTGCAACACTAGCAACACTAGTACTGTAATGATGATTTTG 30878

QY      901 CATTTTTTCATTCAACAATATTTCTGGAACAACCTTATATGCGAGCACTATTTTG 960
DB      30877 CATTTTTTCATTCAACAATATTTCTGGAACAACCTTATATGCGAGCACTATTTTG 30818

QY      961 GAGTCAGAGGATATATATGTTGTAACAAGACAGGAGAAACAAAGCAAGCAACAACA 1020
DB      30817 GAGTCAGAGGATATATATGTTGTAACAAGACAGGAGAAACAAAGCAAGCAACAACA 30758

QY      1021 TCACAGATATATAGACAGATGAGAAAGATTTCAAGTTTATAGTAAATTAACACAC 1080
DB      30757 TCACAGATATATAGACAGATGAGAAAGATTTCAAGTTTATAGTAAATTAACACAC 30698

QY      1081 AAGGCTGAAATGCTGATAGTAAAGGCGTCAAGAAAGCTTCAATGAGAGGTAGCATTT 1140
DB      30697 AAGGCTGAAATGCTGATAGTAAAGGCGTCAAGAAAGCTTCAATGAGAGGTAGCATTT 30638

QY      1141 AAGCAGAGGTGAGCTAGAAATTTGTAATTCAGTTTCAAGTTTATTTGTTGGGTT 1200
DB      30637 AAGCAGAGGTGAGCTAGAAATTTGTAATTTGTAATTTCAAGTTTATTTGTTGGGTT 30578

QY      1201 GGTAAATTAAGCTTTTCCCAAGTGAAGTAACTACAAAGAAAGCTAATTAAGTAG 1260
DB      30577 GGTAAATTAAGCTTTTCCCAAGTGAAGTAACTACAAAGAAAGCTAATTAAGTAG 30518

QY      1261 TGGTGTGCTCTGTGAAAGAGACACTCTGTTTCTGCTCATTTACTGCAACCTTTC 1320
DB      30517 TGGTGTGCTCTGTGAAAGAGACACTCTGTTTCTGCTCATTTACTGCAACCTTTC 30458

QY      1321 ACTTCCAGCACTTTTTCGAAAGCCCTTTCAGAGTCAAGGAGGAGGAGGCTGGCAGT 1380
DB      30457 ACTTCCAGCACTTTTTCGAAAGCCCTTTCAGAGTCAAGGAGGAGGAGGCTGGCAGT 30398

QY      1381 GGGCTTGACATTTTGAACAAGTGAAGCATTTTGTCTCCAGACTCATAGCCCAAGGCT 1440
DB      30397 GGGCTTGACATTTTGAACAAGTGAAGCATTTTGTCTCCAGACTCATAGCCCAAGGCT 30338

QY      1441 AAAGCTGAAGAGGCTTGGGATGCCCCAGAAAGGCCCTGATGAAGCTTGAAGAAAGCTG 1500
DB      30337 AAAGCTGAAGAGGCTTGGGATGCCCCAGAAAGGCCCTGATGAAGCTTGAAGAAAGCTG 30278

```

REFERENCE  
AUTHORS

2 (bases 1 to 112295)  
 Birren, B., Linton, L., Nusbaum, C., Lander, B., Abraham, H., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L.,  
 Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,  
 Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,  
 Graham, L., Grand-Pierre, N., Hagos, B., Heford, A., Horton, L.,  
 Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacroque, K.,  
 Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Lien, C., Liu, G.,  
 Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheeters, R., Meldrim, J., Menues, L., Mihova, T., Mieng, V.,  
 Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,  
 O'Donnell, P., O'Neill, D., Pollara, V., Raymond, C., Rieback, M., Riley, R.,  
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
 Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talama, J., Testaye, S., Theodore, J.,  
 Tirlrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
 Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
 Submitted (25-OCT-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 112295)

REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, B., Ali, A., Allen, N.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,  
 Landers, T., Lehocsky, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marguis, N., Mathews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Menues, L.,  
 Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talama, J., Testaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
 Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 112295)

REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, B., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,  
 Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N.,  
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,  
 Menues, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N.,  
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
 Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Straus, N., Subramanian, A., Talama, J., Testaye, S.,  
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Direct Submission  
 Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 1, 2002 this sequence version replaced g1:19683126.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

## COMMENT

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L10956  
 Center clone name: 2010\_I\_16  
 -----

Only the last 112.3 kb of this clone are being submitted.  
 The remainder overlaps AC068733 [WIGR project L10266].

## FEATURES

source  
 1. 112295  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11"  
 /clone="CTD-2010116"  
 /clone\_1lb="CTDI Human BAC"  
 complement (949..1264)  
 /rpt\_family="AluSq"  
 repeat\_region  
 1269..1568  
 /rpt\_family="L2"  
 complement (2334..2584)  
 /rpt\_family="AluSq"  
 complement (2429..2433)  
 /note="<30 qual SINGL region"  
 complement (2462..2466)  
 /note="<30 qual SINGL region"  
 complement (2478..2503)  
 /note="<30 qual SINGL region"  
 complement (2508..2537)  
 /note="<30 qual SINGL region"  
 complement (2577..2583)  
 /note="<30 qual SINGL region"  
 repeat\_region  
 2772..3185  
 /rpt\_family="L2"  
 complement (4149..4288)  
 /rpt\_family="L2"  
 repeat\_region  
 4268..4343  
 /rpt\_family="L2"  
 repeat\_region  
 4453..4524  
 /rpt\_family="L2"  
 repeat\_region  
 4563..5531  
 /rpt\_family="L2"  
 complement (5560..5857)  
 /rpt\_family="AluSx"  
 5861..6365  
 /rpt\_family="L2"  
 complement (6378..7246)  
 /rpt\_family="L2"  
 repeat\_region  
 7898..7942  
 /rpt\_family="GC\_rich"  
 repeat\_region  
 8243..8264  
 /rpt\_family="GC\_rich"  
 complement (8355..8589)  
 /rpt\_family="MIR"  
 complement (8875..9148)  
 /rpt\_family="AluJo"  
 9264..9349  
 /rpt\_family="L2"  
 complement (9361..9480)  
 /rpt\_family="L1MC5"  
 complement (9584..9639)  
 repeat\_region



Db 173311 ACAACTTTGGCTTGACGTCGACATCATG-TCCTCCACTTTCCTGACTATTCATGATC 173253  
 Qy 8101 TTTAGCTCAGTTCACATGATACCTCCCTGAGTGTCTTTTCTAGTAAGATTC 8160  
 Db 173252 TTTCACTCAGTTCACATGATACCTCCCTGAGTGTCTTTTCTAGTAAGATTC 173193  
 Qy 8161 CAGAGAGGAAATCTGAATGGCCAGTCATATTTTTCAGACACACACATTAAGTGT 8220  
 Db 173192 CAGAGAGGAAATCTGAATGGCCAGTCATATTTTTCAGACACACACATTAAGTGT 173133  
 Qy 8221 GATTCGACCTCTATGATTTGGCTACATTAATGGGTTGGAACTCATCTTACTTCAATG 8280  
 Db 173132 GATTCGACCTCTATGATTTGGCTACATTAATGGGTTGGAACTCATCTTACTTCAATG 173073  
 Qy 8281 CACAAAGCAGCATAGCTGTGTTCTCAAAATAGAGCCCTGGGCGAGGTGGTGGCTCA 8340  
 Db 173072 CACAAAGCAGCATAGCTGTGTTCTCAAAATAGAGCCCTGGGCGAGGTGGTGGCTCA 173013  
 Qy 8341 TGCCTATTAATCCCAACATCTGTGGAGCCGAGGGGGGAGATCATCTTGAATCAGAGTT 8400  
 Db 173012 TGCCTATTAATCCCAACATCTGTGGAGCCGAGGGGGGAGATCATCTTGAATCAGAGTT 172953  
 Qy 8401 CTAGACACCTCTGGGCAATGTGTGAATCTCATCTCTAATAAATAAATAAATTAGC 8460  
 Db 172952 CTAGACACCTCTGGGCAATGTGTGAATCTCATCTCTAATAAATAAATAAATTAGC 172893  
 Qy 8461 CAGGTGTGGTGAGTCACACATGTCACAGCTGTCAAGAGCTGAGGTGGAGGATG 8520  
 Db 172892 CAGGTGTGGTGAGTCACACATGTCACAGCTGTCAAGAGCTGAGGTGGAGGATG 172833  
 Qy 8521 CTCGAGTGTGGAGGAGCAGATTTGAGTAACTGTGCTGTGCAATCAGCTG 8580  
 Db 172832 CTCGAGTGTGGAGGAGCAGATTTGAGTAACTGTGCTGTGCAATCAGCTG 172773  
 Qy 8581 GGTGACATTTGAGACCTGTCTCAAAAACAAATTAATTAATTAATTAATGTTTC 8640  
 Db 172772 GGTGACATTTGAGACCTGTCTCAAAAACAAATTAATTAATTAATTAATGTTTC 172713  
 Qy 8641 CTGAGCAGGTTAATTCAGTGGGAACTCCAGGGGAGGTGATATGTCAGTCACTGCT 8700  
 Db 172712 CTGAGCAGGTTAATTCAGTGGGAACTCCAGGGGAGGTGATATGTCAGTCACTGCT 172653  
 Qy 8701 GTATCTCAGTACACGGCTAATTAAGAACTTGTGTAGCAGCAAGAACATAGTATTT 8760  
 Db 172652 GTATCTCAGTACACGGCTAATTAAGAACTTGTGTAGCAGCAAGAACATAGTATTT 172593  
 Qy 8761 ACTCAACAAATATTTGTGAGCATCTGATPAAGAGTGGCATTTGCTTAGGCACTGAT 8820  
 Db 172592 ACTCAACAAATATTTGTGAGCATCTGATPAAGAGTGGCATTTGCTTAGGCACTGAT 172533  
 Qy 8821 ACAAGTGTCAATGTCAGCAAGAGTCCCTGACAGGCTCTGCTAAGTGAAGAG 8880  
 Db 172532 ACAAGTGTCAATGTCAGCAAGAGTCCCTGACAGGCTCTGCTAAGTGAAGAG 172473  
 Qy 8881 GACAATTAAGAAAGAAAGAAAGAAAGAAATTAATTTAGTAAATTAAGGTTTAA 8940  
 Db 172472 GACAATTAAGAAAGAAAGAAAGAAAGAAATTAATTTAGTAAATTAAGGTTTAA 172413  
 Qy 8941 GAAAAATTAAGACAGATGATGGGATAGAGATGAGAGATGAGGGCTGTCTTGAAGAA 9000  
 Db 172412 GAAAAATTAAGACAGATGATGGGATAGAGATGAGAGATGAGGGCTGTCTTGAAGAA 172353  
 Qy 9001 TGAATTTTGAAGCTGAGACTTTCAGTGTGAAGAAATTAACACAGATGTCTGAGGA 9060  
 Db 172352 TGAATTTTGAAGCTGAGACTTTCAGTGTGAAGAAATTAACACAGATGTCTGAGGA 172293  
 Qy 9061 AAAGCATTTTGAAGAGGAGTGAAGACATCTTCAAGGAATCAAGAGAAAGCTGTG 9120  
 Db 172292 AAAGCATTTTGAAGAGGAGTGAAGACATCTTCAAGGAATCAAGAGAAAGCTGTG 172233  
 Qy 9121 AGGCTGGAACACAGAGAAAGACAGGTGGGTGCTTGAAGGGCAGGGAACGCACTG 9180  
 Db 172232 AGGCTGGAACACAGAGAAAGACAGGTGGGTGCTTGAAGGGCAGGGAACGCACTG 172173

Qy 9181 AGGTTACTAGACCTGTGAAGGTTTCAACCAATAAAGGAGTCAATCAGAAAGCTTGA 9240  
 Db 172172 AGGTTACTAGACCTGTGAAGGTTTCAACCAATAAAGGAGTCAATCAGAAAGCTTGA 172113  
 Qy 9241 GAGGGCTGTGATATATTTTCACTCAATTTTATAAAGATCACTGACTTTTTCAG 9300  
 Db 172112 GAGGGCTGTGATATATTTTCACTCAATTTTATAAAGATCACTGACTTTTTCAG 172053  
 Qy 9301 ACTTAAGTATTAAGATTAACACATGTAAGAAATCAGTACATCCGTGAGTT 9360  
 Db 172052 ACTTAAGTATTAAGATTAACACATGTAAGAAATCAGTACATCCGTGAGTT 171993  
 Qy 9361 GTCCAAATTAAGAGTATGACCGCTTGAGTATGATAGCAGAGAGTGGTGA 9420  
 Db 171992 GTCCAAATTAAGAGTATGACCGCTTGAGTATGATAGCAGAGAGTGGTGA 171933  
 Qy 9421 TCACCATATATATTTTGAAGGTAGACTGACAGCATTAATTAATGACTAATAGGCGC 9480  
 Db 171932 TCACCATATATATTTTGAAGGTAGACTGACAGCATTAATTAATGACTAATAGGCGC 171873  
 Qy 9481 GGTGTGTGGCTTACGCTGTATCTTACACTTTTGGAGGCCAAGGCAATGATCAC 9540  
 Db 171872 GGTGTGTGGCTTACGCTGTATCTTACACTTTTGGAGGCCAAGGCAATGATCAC 171813  
 Qy 9541 TGAGGTCAAGAGTTCAGACACAGCTTGACCAATGATGAACCTGTCTTACTAATAA 9600  
 Db 171812 TGAGGTCAAGAGTTCAGACACAGCTTGACCAATGATGAACCTGTCTTACTAATAA 171753  
 Qy 9601 TACAAATTAATGCTGGAAATGTGTGCAATGCTGTATCTCACTGACTTGGAGGCTGA 9660  
 Db 171752 TACAAATTAATGCTGGAAATGTGTGCAATGCTGTATCTCACTGACTTGGAGGCTGA 171693  
 Qy 9661 GGCAGAGAAATGCTTGAACCTGAGAGTGAATGTTGACGTGAGCCGAAATTCACACT 9720  
 Db 171692 GGCAGAGAAATGCTTGAACCTGAGAGTGAATGTTGACGTGAGCCGAAATTCACACT 171633  
 Qy 9721 GCACTTCAGCTGGGGAACAAGTGAACCTCGCTCTTAATTAATGAATGAATG 9780  
 Db 171632 GCACTTCAGCTGGGGAACAAGTGAACCTCGCTCTTAATTAATGAATGAATG 171573  
 Qy 9781 ATATCAGTCAGATGAGGAGGAAAGGAAAGGCTTCAAGATGACTCAGCTTTCGTGACT 9840  
 Db 171572 ATATCAGTCAGATGAGGAGGAAAGGAAAGGCTTCAAGATGACTCAGCTTTCGTGACT 171513  
 Qy 9841 CAGCACTGAGTGGCTGTGTTTCTTAAATTTGGAAAGACTAGGAGTGTG 9900  
 Db 171512 CAGCACTGAGTGGCTGTGTTTCTTAAATTTGGAAAGACTAGGAGTGTG 171453  
 Qy 9901 TGTGTGTGGGGGCAAAATCAGTTTGGCATTAATGTTTGGGCTATTTGGCACC 9960  
 Db 171452 TGTGTGTGGGGGCAAAATCAGTTTGGCATTAATGTTTGGGCTATTTGGCACC 171393  
 Qy 9961 CATTAAGCATGTGAGTGGCAGCTGATTTGAGGCTTAAC 10000  
 Db 171392 CATTAAGCATGTGAGTGGCAGCTGATTTGAGGCTTAAC 171353

RESULT 2  
 AC084337/c 112295 bp DNA linear pri 01-JUN-2002  
 LOCUS AC084337 Homo sapiens chromosome 11, clone CTD-2010116, complete sequence.  
 DEFINITION AC084337  
 ACCESSION AC084337.7 GI:21306722  
 VERSION AC084337.7  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 112295)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, B.  
 TITLE Homo sapiens chromosome 11, clone CTD-2010116  
 JOURNAL Unpublished

QY 5881 CCATGTCGCCACCTATGCTTCAGTGTGAGATGCCCCCACTCCCAATGTGCTCT 5940  
DB 175471 CCATGTCGCCACCTATGCTTCAGTGTGAGATGCCCCCACTCCCAATGTGCTCT 175412  
QY 5941 CACATCTCTTTTACTTTATCTTCCCATCTTGGACATTTTCTCATTTGTTCATCACTGTGT 6000  
DB 175411 CACATCTCTTTTACTTTATCTTCCCATCTTGGACATTTTCTCATTTGTTCATCACTGTGT 175352  
QY 6001 CACTTATTTTGTGCTCCCTGCTCCCATCTTCTGACATGCTCTGTCATCCCTCATCTCT 6060  
DB 175351 CACTTATTTTGTGCTCCCTGCTCCCATCTTCTGACATGCTCTGTCATCCCTCATCTCTCT 175292  
QY 6061 GAGGCATATTTTCTCAATCTTGTCTGTCAAGGCCCAAGCCCTTAATCTTCACTCTGTCT 6120  
DB 175291 GAGGCATATTTTCTCAATCTTGTCTGTCAAGGCCCAAGCCCTTAATCTTCACTCTGTCT 175232  
QY 6121 ACCATCTACTCCCATGCTGTGTGCTCCCTGTGACCTCTCTGAGGCCCTCATGACTCTTGT 6180  
DB 175231 ACCATCTACTCCCATGCTGTGTGCTCCCTGTGACCTCTCTGAGGCCCTCATGACTCTTGT 175172  
QY 6181 GTTCTCTGCTCAATGCTCCCTGTGAGGCCCTGTGAGGCCCTCTGCTCCCTGAGACCTCTTA 6240  
DB 175171 GTTCTCTGCTCAATGCTCCCTGTGAGGCCCTGTGAGGCCCTCTGCTCCCTGAGACCTCTTA 175112  
QY 6241 TGTGTCTGTACT 6300  
DB 175111 TGTGTCTGTACT 175052  
QY 6301 CCCCCCTGATTTTCTCTGACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6360  
DB 175051 CCCCCCTGATTTTCTCTGACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 174992  
QY 6361 CCCCCCTGATTTTCTCTGACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6420  
DB 174991 CCCCCCTGATTTTCTCTGACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 174932  
QY 6421 TGTGTGTGTACTTTAGACCTTTAGAGCTGTGCTTCACTAGCTCTGTGAGGTTTCTCTCT 6480  
DB 174931 TGTGTGTGTACTTTAGACCTTTAGAGCTGTGCTTCACTAGCTCTGTGAGGTTTCTCTCT 174872  
QY 6481 AGTAGCCATGAGATATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6540  
DB 174871 AGTAGCCATGAGATATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 174812  
QY 6541 TATATCTTGAAGGCTGTCTGAGTTCACTGAGGAGTAATGATGATGCTGCAAGTTGAG 6600  
DB 174811 TATATCTTGAAGGCTGTCTGAGTTCACTGAGGAGTAATGATGATGCTGCAAGTTGAG 174752  
QY 6601 AATATCCCCAAGAGAGTGAAGCAACATCATCTCCATCTGAGATATATGTATATTA 6660  
DB 174751 AATATCCCCAAGAGAGTGAAGCAACATCATCTCCATCTGAGATATATGTATATTA 174692  
QY 6661 GGAAGCTGTGATGATATATAATGTGAAAAATTTTTTTCACACAAAATTTTTTGGCT 6720  
DB 174691 GGAAGCTGTGATGATATATAATGTGAAAAATTTTTTTCACACAAAATTTTTTGGCT 174632  
QY 6721 CCGAGACCTTTGACAAATTTGACAGTTAGATATCAAGTTCTGTTGAAAAATATACATCA 6780  
DB 174631 CCGAGACCTTTGACAAATTTGACAGTTAGATATCAAGTTCTGTTGAAAAATATACATCA 174572  
QY 6781 CCACATGAGAGCAAAATCTCCACAGCAGATTTGCACTATTAATTAAGAACATACAGCTTA 6840  
DB 174571 CCACATGAGAGCAAAATCTCCACAGCAGATTTGCACTATTAATTAAGAACATACAGCTTA 174512  
QY 6841 GATGAAACACACCTGTGATGAAATATCAACATTTAACTGAGAACATACGCAATAGTAA 6900  
DB 174511 GATGAAACACACCTGTGATGAAATATCAACATTTAACTGAGAACATACGCAATAGTAA 174452  
QY 6901 GAACACATTAATGATCAAGAGACACACAGGCATGTGTGGAGCCCATTTGGAGAGACACACA 6960  
DB 174451 GAACACATTAATGATCAAGAGACACACAGGCATGTGTGGAGCCCATTTGGAGAGACACACA 174392  
QY 6961 GACAAAGTGAATGACAAAAGAGAGAGTGAAGATTTGTGAAAAACAGGCGCAC 7020

DB 174391 GACAAAGTGAATGACAAAAGAGAGAGAGTGAAGATTTGTGAAAAACAGGCGCAC 174332  
QY 7021 AGGAAACACACGAAATTAAGAGAGACACAGGCCATCTAGAGATACAGAACTTCATGG 7080  
DB 174331 AGGAAACACACGAAATTAAGAGAGACACAGGCCATCTAGAGATACAGAACTTCATGG 174272  
QY 7081 CCATGTGACCATTAATGAGAAATCTACTGAATCTTAATGAAAAATGTCATGATGTCTC 7140  
DB 174271 CCATGTGACCATTAATGAGAAATCTACTGAATCTTAATGAAAAATGTCATGATGTCTC 174212  
QY 7141 ATAGCTGTTGAGAGAGCCCAACAGATGAGAGAACCTTATATTAATAATCCAGGCC 7200  
DB 174211 ATAGCTGTTGAGAGAGCCCAACAGATGAGAGAACCTTATATTAATAATCCAGGCC 174152  
QY 7201 GGGCGGTGTGAGTCAAGCTCTTAATCTTGAACCTTTGGAGGCTGAGGAGGTGATTC 7260  
DB 174151 GGGCGGTGTGAGTCAAGCTCTTAATCTTGAACCTTTGGAGGCTGAGGAGGTGATTC 174092  
QY 7261 TTGAGCGGCTTGAGGCTTGAAGATTTGAGACAGGCTGAGCAACATGSCAAAACCTCATCT 7320  
DB 174091 TTGAGCGGCTTGAGGCTTGAAGATTTGAGACAGGCTGAGCAACATGSCAAAACCTCATCT 174032  
QY 7321 CTACAAAATAATTAATAATTAAGTGGGTGTGAGTGTGCTCTATAGTCCATCTACTT 7380  
DB 174031 CTACAAAATAATTAATAATTAAGTGGGTGTGAGTGTGCTCTATAGTCCATCTACTT 173972  
QY 7381 CAGAGCTGAGCCCGGAGAGTGTGAGGCTTCACTGAGGCTGTATGCTACTGACATCCA 7440  
DB 173971 CAGAGCTGAGCCCGGAGAGTGTGAGGCTTCACTGAGGCTGTATGCTACTGACATCCA 173912  
QY 7441 GCGTGGGTGACAGAGTGAAGACCATGTCTCAAAAAAACAACAAAAACAACAAAAACA 7500  
DB 173911 GCGTGGGTGACAGAGTGAAGACCATGTCTCAAAAAAACAACAAAAACAACAAAAACA 173852  
QY 7501 CAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7560  
DB 173851 CAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 173792  
QY 7561 CTAATTTGGTAAATATCAACAAATGACCAAGCTGACAGATGAGGAGTCCCTCCCTCTC 7620  
DB 173791 CTAATTTGGTAAATATCAACAAATGACCAAGCTGACAGATGAGGAGTCCCTCCCTCTC 173732  
QY 7621 ACTGTAATTTTCTTCTCTGACCTCAGATTTTGTGTGTGTGTGTGTGTGTGTGTGTG 7680  
DB 173731 ACTGTAATTTTCTTCTCTGACCTCAGATTTTGTGTGTGTGTGTGTGTGTGTGTGTG 173672  
QY 7681 ATGAGCTCTCACTGTGTCAACCAAGCTGAGTGAATGAGCAATGCTGTGTGTGTGTGTG 7740  
DB 173671 ATGAGCTCTCACTGTGTCAACCAAGCTGAGTGAATGAGCAATGCTGTGTGTGTGTGTG 173612  
QY 7741 CCTGTGCTCTG 7800  
DB 173611 CCTGTGCTCTG 173552  
QY 7801 AGGCGCATCAACCAATGCTGCTAATTTTGTATTTTGTGTTTGTGTTTGTGTTTGTGTTT 7860  
DB 173551 AGGCGCATCAACCAATGCTGCTAATTTTGTATTTTGTGTTTGTGTTTGTGTTTGTGTTT 173492  
QY 7861 TTAATTAATTAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTAT 7920  
DB 173491 TTAATTAATTAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTAT 173432  
QY 7921 GTCTGCAATCTCTGACCTCAGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7980  
DB 173431 GTCTGCAATCTCTGACCTCAGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 173372  
QY 7981 GAGGCATGAGCAACACAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 8040  
DB 173371 GAGGCATGAGCAACACAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 173312  
QY 8041 ACAACTTGGCTTGAACGTGACCATCATGTTCTCTCACTTGTCTCACTTGTCTCATGATC 8100

Db 177691 GGTATCTTCTCAGGTTTATCTCCCTCCCTGCACATTCAGTGTGCCAGCATACGAA 177632  
Qy 3721 TCTACATGAGGTTGAGAGCACTGCTTCCATGTTTGGGCTGTGATGTGCTCCCTCT 3780  
Db 177631 TCTACATGAGGTTGAGAGCACTGCTTCCATGTTTGGGCTGTGATGTGCTCCCTCT 177572  
Qy 3781 GCTGTGATACCCCTTCTCTCACTTGTCAACCTTGAAAATTCCTGTGATTTTCACTCT 3840  
Db 177571 GCTGTGATACCCCTTCTCTCACTTGTCAACCTTGAAAATTCCTGTGATTTTCACTCT 177512  
Qy 3841 TGGGCCCAATGCTCTCTCTTGGTGTGAACCTTCCAACTTCTCTAGGCACTTAAAG 3900  
Db 177511 TGGGCCCAATGCTCTCTCTTGGTGTGAACCTTCCAACTTCTCTAGGCACTTAAAG 177452  
Qy 3901 CACTGTCTATATCTCAGTGTGACTTCTTACCTTACCTTGGTGTGAGGCTAGG 3960  
Db 177451 CACTGTCTATATCTCAGTGTGACTTCTTACCTTGGTGTGAGGCTAGG 177392  
Qy 3961 ATTGCAGAGTCTTCTCTCTTGTGTAGTGAACCTTCTGAAGTGAAGACAGAGCTT 4020  
Db 177391 ATTGCAGAGTCTTCTCTCTTGTGTAGTGAACCTTCTGAAGTGAAGACAGAGCTT 177332  
Qy 4021 GTTATCTCAGTGTCTCTCAACATGCTGATATAGTATTAATTCAGTACTGTTCTT 4080  
Db 177331 GTTATCTCAGTGTCTCTCAACATGCTGATATAGTATTAATTCAGTACTGTTCTT 177272  
Qy 4081 GGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 4140  
Db 177271 GGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 177212  
Qy 4141 ATTGCAGATGAACAGTGTGTGATGTTTGTCACTGATGATGTGAGATTAATGT 4200  
Db 177211 ATTGCAGATGAACAGTGTGTGATGTTTGTCACTGATGATGTGAGATTAATGT 177152  
Qy 4201 GCCACAGAGTGTCTGGGTACAGAGCTAGAGCATGTGTATTAATGAATGACTGGA 4260  
Db 177151 GCCACAGAGTGTCTGGGTACAGAGCTAGAGCATGTGTATTAATGAATGACTGGA 177092  
Qy 4261 TTTTGCACAACTGAGAGTGTGTAAATGTCAGAAAGACAGACATTTTGTCCACAAATG 4320  
Db 177091 TTTTGCACAACTGAGAGTGTGTAAATGTCAGAAAGACAGACATTTTGTCCACAAATG 177032  
Qy 4321 ACTGAGATGTGTAGGGCCACAGAAAGATATGATTAAGACAGTATTAATTAATGTGT 4380  
Db 177031 ACTGAGATGTGTAGGGCCACAGAAAGATATGATTAAGACAGTATTAATTAATGTGT 176972  
Qy 4381 GTTAATGCAAGTGTGAGTATCTGGGATGACAGTCAAAAAGAGATCTTTGAATGC 4440  
Db 176971 GTTAATGCAAGTGTGAGTATCTGGGATGACAGTCAAAAAGAGATCTTTGAATGC 176912  
Qy 4441 AGGGGACAAAGTCTGGGATATACCTCTCTGAAAAGAGAAAGATACCAAAGTTGC 4500  
Db 176911 AGGGGACAAAGTCTGGGATATACCTCTCTGAAAAGAGAAAGATACCAAAGTTGC 176852  
Qy 4501 TCCAAAGATGAATTTCTCTGAATCCCATCCCACTGATGAGCTGTGATGTCACGCTG 4560  
Db 176851 TCCAAAGATGAATTTCTCTGAATCCCATCCCACTGATGAGCTGTGATGTCACGCTG 176792  
Qy 4561 GAGAATGTCAAGTGAAGGCTCTCTTCTTCCAGAGTCAATCCAGCTGATATCCAGAA 4620  
Db 176791 GAGAATGTCAAGTGAAGGCTCTCTTCTTCCAGAGTCAATCCAGCTGATATCCAGAA 176732  
Qy 4621 ACCCTGAGATAGTGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
Db 176731 ACCCTGAGATAGTGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 176672  
Qy 4681 GTGGTGTGAGGCACTATGTGGCTTTTAAAGAAAGGAATGAAGAGCAAGTATGAT 4740  
Db 176671 GTGGTGTGAGGCACTATGTGGCTTTTAAAGAAAGGAATGAAGAGCAAGTATGAT 176612  
Qy 4741 CTAGAGTCACTGAGGAGAGAGAGAGATCTTCTGTGTGAAGGCACTGTGTATAT 4800  
Db 176611 CTAGAGTCACTGAGGAGAGAGAGAGATCTTCTGTGTGAAGGCACTGTGTATAT 176552

Qy 4801 CTGTGTGAATAGTCCCTCATTTGTTCCCATGACCTTAAAGACAAATCTACCTCTTATG 4860  
Db 176551 CTGTGTGAATAGTCCCTCATTTGTTCCCATGACCTTAAAGACAAATCTACCTCTTATG 176492  
Qy 4861 CTTCATTAAGTCTCTCAATGGCCAAATCCCTATTTGGCCCTTCACTTATTAAT 4920  
Db 176491 CTTCATTAAGTCTCTCAATGGCCAAATCCCTATTTGGCCCTTCACTTATTAAT 176432  
Qy 4921 ATACTTTTACCTTAACATTAAGTCCAGAAACCTTAATGCTATCTGTATCACTGCTT 4980  
Db 176431 ATACTTTTACCTTAACATTAAGTCCAGAAACCTTAATGCTATCTGTATCACTGCTT 176372  
Qy 4981 GCTTCAGTGTGAATCTTCTCTCTGGGGGTCCATCTCTGTTGAGCTTTTA 5040  
Db 176371 GCTTCAGTGTGAATCTTCTCTCTGGGGGTCCATCTCTGTTGAGCTTTTA 176312  
Qy 5041 ATTCTTACTTCAAGATTTCACTTAATATCATCTTCCCTGGAGATTTTCCAGACTCT 5100  
Db 176311 ATTCTTACTTCAAGATTTCACTTAATATCATCTTCCCTGGAGATTTTCCAGACTCT 176252  
Qy 5101 CCCACTGCTTGTGAGCTGATCTGTGTGTTTGTCTGTGATTTGTGTATGATCA 5160  
Db 176251 CCCACTGCTTGTGAGCTGATCTGTGTGTTTGTCTGTGATTTGTGTATGATCA 176192  
Qy 5161 CCTCTCTTACGCTCTCTGATGGCTGTGAGCTCATGTGTGATCAATATCTGG 5220  
Db 176191 CCTCTCTTACGCTCTCTGATGGCTGTGAGCTCATGTGTGATCAATATCTGG 176132  
Qy 5221 CCCATCTGGGACCCAGAGAAAGCAAGAGAGGGGTAAACCGGTCTCAACCAATGCT 5280  
Db 176131 CCCATCTGGGACCCAGAGAAAGCAAGAGAGGGGTAAACCGGTCTCAACCAATGCT 176072  
Qy 5281 GTTGATTAATGAACAAAGTGAACCGAGTGTCTTGGGACTTGGCTACGGGAAACATG 5340  
Db 176071 GTTGATTAATGAACAAAGTGAACCGAGTGTCTTGGGACTTGGCTACGGGAAACATG 176012  
Qy 5341 AAGGAGCTTCTGAGCAGCTGTGGGAACTGTCTCTGCTGCTGAGATGGCTGGCCGC 5400  
Db 176011 AAGGAGCTTCTGAGCAGCTGTGGGAACTGTCTCTGCTGCTGAGATGGCTGGCCGC 175952  
Qy 5401 TACTACTGCTTCAGGGATTAACCAATCTCTGCTTGTGACCTGTCAAGGGAGAGTGCCT 5460  
Db 175951 TACTACTGCTTCAGGGATTAACCAATCTCTGCTTGTGACCTGTCAAGGGAGAGTGCCT 175892  
Qy 5461 CCCAGGTACCCGCGGATGTCCGAGACTTCTTCAATGCTGCTGCTGCGAGAGTGAAGA 5520  
Db 175891 CCCAGGTACCCGCGGATGTCCGAGACTTCTTCAATGCTGCTGCGAGAGTGAAGA 175832  
Qy 5521 GCCCTAGCACTTGAACCTGTGAAATTCATCACTTCCCTGAGCTGTGATCTCAG 5580  
Db 175831 GCCCTAGCACTTGAACCTGTGAAATTCATCACTTCCCTGAGCTGTGATCTCAG 175772  
Qy 5581 TGTCTAGCTCTCACTTAACTCCGTGTGCAACCTTGTGACCTTAACTAGCCCAT 5640  
Db 175771 TGTCTAGCTCTCACTTAACTCCGTGTGCAACCTTGTGACCTTAACTAGCCCAT 175712  
Qy 5641 TCCATTTCTGAATTTTCCATTTGCTTCCATATGGGAAACCCACACCCCACTAACCCAC 5700  
Db 175711 TCCATTTCTGAATTTTCCATTTGCTTCCATATGGGAAACCCACACCCCACTAACCCAC 175652  
Qy 5701 CATCTCTTCACTTGAACCTCACTCTGACCTTGGCCCTCTTGTGTCTCTCAACC 5760  
Db 175651 CATCTCTTCACTTGAACCTCACTCTGACCTTGGCCCTCTTGTGTCTCTCAACC 175592  
Qy 5761 ATTTCTCTTCCAGGCACTGACACAGAAATGGGATGCGCATGGGAAAGATACCA 5820  
Db 175591 ATTTCTCTTCCAGGCACTGACACAGAAATGGGATGCGCATGGGAAAGATACCA 175532  
Qy 5821 TGGCCCTGAGATATGCGGTGAGCCCACTATGATCTTGTCTGCACTGACGTCTGACAA 5880  
Db 175531 TGGCCCTGAGATATGCGGTGAGCCCACTATGATCTTGTCTGCACTGACGTCTGACAA 175472

```
QY 1501 TTCTGAGATTTCTAAGAGTTTATCNGTNGTGTGTTACTAAAGTAAAGTAT 1560
Db 179851 TTCTGAGATTTCTAAGATTTATCTGTGTGTGTGTATCTAAAGTAAAGTAT 179792
QY 1561 TGTGTCTGTAGTGTCTTGAAGACAGAGGCTTGAACAAGTACACAGCAATATAGTCCCT 1620
Db 179791 TGTGTCTGTAGTGTCTTGAAGACAGAGGCTTGAACAAGTACACAGCAATATAGTCCCT 179732
QY 1621 CCTTTTCTACCTCCCCCATTTGTGAGATTAATCAATCAACAAAGGTATCTCAGTCT 1680
Db 179731 CCTTTTCTACCTCCCCCATTTGTGAGATTAATCAATCAACAAAGGTATCTCAGTCT 179672
QY 1681 ACTCATCTCCCTGATTTATGAGATGCTGAGCCCATTTGCCAGTGTGAAGTACAGCTGGA 1740
Db 179671 ACTCATCTCCCTGATTTATGAGATGCTGAGCCCATTTGCCAGTGTGAAGTACAGCTGGA 179612
QY 1741 CGTCAGCAGTGTAGCCCAAGTTACTGTCTGTAATAATGCTGAAGGGGTGTGGGGGACAGCTG 1800
Db 179611 CGTCAGCAGTGTAGCCCAAGTTACTGTGTAAATGCTGAAGGGGTGTGGGGGACAGCTG 179552
QY 1801 CCGGGAAAAAGAGTCTTGATTCAGATTTCTGTCCAGACCTTGACCTTATTTGCAGTGA 1860
Db 179551 CCGGGAAAAAGAGTCTTGATTCAGATTTCTGTCCAGACCTTGACCTTATTTGCAGTGA 179492
QY 1861 TGTATCAGCCATATTGGCTTAGTCTGCTGGGAGACAGCAATCCAGTAGAGTTGAGG 1920
Db 179491 TGTATCAGCCATATTGGCTTAGTCTGCTGGGAGACAGCAATCCAGTAGAGTTGAGG 179432
QY 1921 TGGGGGTGTGTCTGTCTGCCAATCTATATAGGAGTTCACCTGTACCCAGAGCTGTCC 1980
Db 179431 TGGGGGTGTGTCTGTCTGCCAATCTATATAGGAGTTCACCTGTACCCAGAGCTGTCC 179372
QY 1981 TGTGGCTCTGTGACCTTCAGATGAGTCTAGGTTACTGGAGACACCCGTTGCACTGGGGTGT 2040
Db 179371 TGTGGCTCTGTGACCTTCAGATGAGTCTAGGTTACTGGAGACACCCGTTGCACTGGGGTGT 179312
QY 2041 GGAGCCATGCTGTCTGTGSCCATTTGCCACCCCTCTCTCCGTGAGTAAAGTGGAG 2100
Db 179311 GGAGCCATGCTGTCTGTGSCCATTTGCCACCCCTCTCTCCGTGAGTAAAGTGGAG 179252
QY 2101 TGAAGCGAAGATTGATTTCTGGGCTAGGGTAAAGTAAAGGCGCAGTTTAAAGCCTCGAT 2160
Db 179251 TGAAGCGAAGATTGATTTCTGGGCTAGGGTAAAGTAAAGGCGCAGTTTAAAGCCTCGAT 179192
QY 2161 CAAATTTGGGGTCAAGGGGCTATAGGAAAGGATCGTCCCAATGATCAAGATATCTATT 2220
Db 179191 CAAATTTGGGGTCAAGGGGCTATAGGAAAGGATCGTCCCAATGATCAAGATATCTATT 179132
QY 2221 TTGTTCTCCCTAGAGCTAGTGGCCATGGAATGTTGCTGAAGGCGAGACCAAGCAAGACC 2280
Db 179131 TTGTTCTCCCTAGAGCTAGTGGCCATGGAATGTTGCTGAAGGCGAGACCAAGCAAGACC 179072
QY 2281 CAGACGTGATGTGTAGGACCCTGATCTCCCTAAGTCTGTCTTATCTGTCTGTGTCTC 2340
Db 179071 CAGACGTGATGTGTAGGACCCTGATCTCCCTAAGTCTGTCTTATCTGTCTGTGTCTC 179012
QY 2341 TGCATTTTATCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400
Db 179011 TGCATTTTATCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 178952
QY 2401 ACCGTCAACCCAGAAAGCTGTGATGAGTGGAGCTTATGATGATCAACCCGTGAGTAC 2460
Db 178951 ACCGTCAACCCAGAAAGCTGTGATGAGTGGAGCTTATGATGATCAACCCGTGAGTAC 178892
QY 2461 AATGGAACCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
Db 178891 AATGGAACCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 178832
QY 2521 AGACTTACTCTCTTCACAAAGGATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2580
Db 178831 AGACTTACTCTCTTCACAAAGGATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 178772
QY 2581 TGGTATCACTGACAGCTTCTCTCAAGCTATCTCAGTAGAGTCAAAAGTTTCTCACTGGGC 2640
Db 178771 TGGTATCACTGACAGCTTCTCTCAAGCTATCTCAGTAGAGTCAAAAGTTTCTCACTGGGC 178712
QY 2641 CCTCAGTAGTGTGGTTTTTCAAGGAGATTGTGTGTGAAGATCAAAATGGACCCG 2700
Db 178711 CCTCAGTAGTGTGGTTTTTCAAGGAGATTGTGTGTGAAGATCAAAATGGACCCG 178652
QY 2701 GAGTTAATCTCAGAGAGATGGAAGAAATTTCCCAAGCCCTGTGAGTGTCTCAATCCGTCA 2760
Db 178651 GAGTTAATCTCAGAGAGATGGAAGAAATTTCCCAAGCCCTGTGAGTGTCTCAATCCGTCA 178592
QY 2761 GGTCAACAAGTGTCTTCTGTATCAAGTACTGTGTGGCCAAATATCAGGCGCAGCTGGA 2820
Db 178591 GGTCAACAAGTGTCTTCTGTATCAAGTACTGTGTGGCCAAATATCAGGCGCAGCTGGA 178532
QY 2821 AAGGGCTGGAATCGACACTGGGGACCTTCCCAATGAGCTTGTGATGAGCCCATAG 2880
Db 178531 AAGGGCTGGAATCGACACTGGGGACCTTCCCAATGAGCTTGTGATGAGCCCATAG 178472
QY 2881 CAATAGTAGAGATTTCTTCCATGAGCCCTTCTGTAAAGCTTGGGCTAAGG 2940
Db 178471 CAATAGTAGAGATTTCTTCCATGAGCCCTTCTGTAAAGCTTGGGCTAAGG 178412
QY 2941 GAGTGTGATGCTGTGGGCTGTGGCAGTGTGACCAATCCAGTGTGTCTTCACTAG 3000
Db 178411 GAGTGTGATGCTGTGGGCTGTGGCAGTGTGACCAATCCAGTGTGTCTTCACTAG 178352
QY 3001 TCTTATGTTTACACCGCTGTCTGTATCTCAACTGTGAGCCATCTTTTCTCTGAC 3060
Db 178351 TCTTATGTTTACACCGCTGTCTGTATCTCAACTGTGAGCCATCTTTTCTCTGAC 178292
QY 3061 AATTGCTTCCCTTCCCTTCAAGACCTGTATCAATCCCTTCAATAGGCAAGGCAAGT 3120
Db 178291 AATTGCTTCCCTTCCCTTCAAGACCTGTATCAATCCCTTCAATAGGCAAGGCAAGT 178232
QY 3121 TGTTCACCAACAATGGACCTGTGATGAAGAGATGTGGAGCATGATGATGATGATGATGAT 3180
Db 178231 TGTTCACCAACAATGGACCTGTGATGAAGAGATGTGGAGCATGATGATGATGATGATGAT 178172
QY 3181 GTGTGACCAAGTCCCATGTACCCAGGCTGTGTCTGTCAAGAGAAAGGGGTGCTTTTC 3240
Db 178171 GTGTGACCAAGTCCCATGTACCCAGGCTGTGTCTGTCAAGAGAAAGGGGTGCTTTTC 178112
QY 3241 CTACCTTGTGCAAAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3300
Db 178111 CTACCTTGTGCAAAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 178052
QY 3301 CTGTCTCTTTTCTCTCTCTCAATGTCTCAATCTCAATCTCAATCTCAATCTCAATCT 3360
Db 178051 CTGTCTCTTTTCTCTCTCTCAATGTCTCAATCTCAATCTCAATCTCAATCTCAATCT 177992
QY 3361 TCTTGGGGAATCATCTAGCTTTCATATACCTTCCCTCAGTATCTCAACCATACC 3420
Db 177991 TCTTGGGGAATCATCTAGCTTTCATATACCTTCCCTCAGTATCTCAACCATACC 177932
QY 3421 TTGGCTCTGAGAAATATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3480
Db 177931 TTGGCTCTGAGAAATATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 177872
QY 3481 GACATCTCTTTCAAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3540
Db 177871 GACATCTCTTTCAAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 177812
QY 3541 GCTTTCATCTTCTTCAAAACCAACCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3600
Db 177811 GCTTTCATCTTCTTCAAAACCAACCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 177752
QY 3601 CCAAAATGTGCTGTATCTAACAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3660
Db 177751 CCAAAATGTGCTGTATCTAACAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 177692
QY 3661 GGTATCTTCTCAGGTTTATCTCTCCCTGTGACATTCAGTGTGTGTGTGTGTGTGTGTGTGT 3720
```

```

unsure /note=<30 qual SNGL region"
14582. .14588
/note=<30 qual SNGL region"
unsure 14899. 14904
/note=<30 qual SNGL region"
repeat_region 16077. .16386
/rpc_family="Alusx"
repeat_region 16387. .16988
/rpc_family="L1MA3"
complement(17002. .17075)
/rpc_family="MIR"
repeat_region complement(17907. .18207)
/rpc_family="Alusx"
repeat_region complement(18218. .18368)
/rpc_family="MIR"
repeat_region 18430. .18711
/rpc_family="Alusx"
repeat_region 18712. .18743
/rpc_family="AT_rich"
repeat_region 19184. .19685
/rpc_family="L2"
complement(19738. .19892)
/rpc_family="L2"
repeat_region 19894. .20155
/rpc_family="Alusx"
repeat_region 20167. .20189
/rpc_family="(CAAA)n"
complement(20649. .20961)
repeat_region 21292. .21411
/rpc_family="Alusq"
repeat_region 21454. .21594
/rpc_family="L2"
repeat_region 21774. .22026
/rpc_family="MLTIE"
repeat_region 22048. .22570
/rpc_family="MER41A"
repeat_region 22571. .23617
/rpc_family="MER41A-int"
repeat_region 23618. .23651
/rpc_family="(TG)n"

Query Match 99.8%; Score 9983.2; DB 9; Length 191656;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9996; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 TCCCTCTCCCGAGCGCCAGCAAAATCTGTAGATTCAAGAGGTTCTGACAGCTG 60
DB 181351 TCCCTCTCCCGAGCGCCAGCAAAATCTGTAGATTCAAGAGGTTCTGACAGCTG 181292
QY 61 AAGACAAGTTGTTGAGAAATCTGTAGAGATCATGCGGTCACAGAGGGAATA 120
DB 181291 AAGACAAGTTGTTGAGAAATCTGTAGAGATCATGCGGTCACAGAGGGAATA 181232
QY 121 TAAGTTTCAAGGCTGAGAGGGAAGAAAGGTGAGGGGAGTCTTAGAATAGTGCTC 180
DB 181231 TAAGTTTCAAGGCTGAGAGGGAAGAAAGGTGAGGGGAGTCTTAGAATAGTGCTC 181172
QY 181 CCATTGCCCAACACCCAGAAAGAAAGATGCTCCTGCAATGCGGGAAGGTAGATAGAA 240
DB 181171 CCATTGCCCAACACCCAGAAAGAAAGATGCTCCTGCAATGCGGGAAGGTAGATAGAA 181112
QY 241 CATTGCTGTAGAGAGATGCGATTTGCCAGGCTGCAAGANTCAGAGAGTCCACCTT 300
DB 181111 CATTGCTGTAGAGAGATGCGATTTGCCAGGCTGCAAGANTCAGAGAGTCCACCTT 181052
QY 301 GCCCACTGACTATGAGAGGGAATGATGTTCAAGACATTTTCAATTGTAAGTCAGAA 360
DB 181051 GCCCACTGACTATGAGAGGGAATGATGTTCAAGACATTTTCAATTGTAAGTCAGAA 180992
QY 361 GAGGACATTGAGCTGATGCGCAGAGGCTGTGTGACATGTTGTTCCAGAGGTTCCGAAATG 420
DB 180991 GAGGACATTGAGCTGATGCGCAGAGGCTGTGTGACATGTTGTTCCAGAGGTTCCGAAATG 180932

```

```

QY 421 TGTGTTTCTGTGGAAGAACTTCGAGAGTAGAAAAAGGATCTGAGACTTTTGTA 480
DB 180931 TGTGTTTCTGTGGAAGAACTTCGAGAGTAGAAAAAGGATCTGAGACTTTTGTA 180872
QY 481 AGATTATATATGGAAGCTGTGAGGGGTCTGAGGCATCTGTAGAGATCAAGGCCCTTTCA 540
DB 180871 AGATTATATATGGAAGCTGTGAGGGGTCTGAGGCATCTGTAGAGATCAAGGCCCTTTCA 180812
QY 541 GCGTTGGCTTAGGAGAGAGGGGTCTTGGAATCTCATCTGCCCCATGAGCTGCCCA 600
DB 180811 GCGTTGGCTTAGGAGAGAGGGGTCTTGGAATCTCATCTGCCCCATGAGCTGCCCA 180752
QY 601 TAAATCTTTCTGACTCACTAGGCAATCTCAACAGAAATGGGGAGCTTTGGAGTGG 660
DB 180751 TAAATCTTTCTGACTCACTAGGCAATCTCAACAGAAATGGGGAGCTTTGGAGTGG 180692
QY 661 GCCCAGAAATATCTAGATAGAGATGAGATCCAGAAAGATGATGTGGGCTGAG 720
DB 180691 GCCCAGAAATATCTAGATAGAGATGAGATCCAGAAAGATGATGTGGGCTGAG 180632
QY 721 AACTGGAAGAGAAACAGAGCTGTCAATAAAGGCGTCTGTGACTCTTAGTCTCAT 780
DB 180631 AACTGGAAGAGAAACAGAGCTGTCAATAAAGGCGTCTGTGACTCTTAGTCTCAT 180572
QY 781 ATGCTACTACCTAACCTAACCCCAATCTTAATTTCTCTAACCCTAGAGGGGGGAA 840
DB 180571 ATGCTACTACCTAACCTAACCCCAATCTTAATTTCTCTAACCCTAGAGGGGGGAA 180512
QY 841 ATTGTGAAATTTGGCTGCAACACTAGCAACTAGTCACTTGAATGCAATTTTGG 900
DB 180511 ATTGTGAAATTTGGCTGCAACACTAGCAACTAGTCACTTGAATGCAATTTTGG 180452
QY 901 CATTGTTTTCATTCAACAAATATTTTGGAAACAACCTTATATGCCAGGCACTAATTTTGG 960
DB 180451 CATTGTTTTCATTCAACAAATATTTTGGAAACAACCTTATATGCCAGGCACTAATTTTGG 180392
QY 961 GAGTCAGGATATATATATGTTAAACAAGCAGGCAAAACAAGCAAAAGCAACAACCA 1020
DB 180391 GAGTCAGGATATATATATGTTAAACAAGCAGGCAAAACAAGCAAAAGCAACAACCA 180332
QY 1021 TCACGATATAGTAGACAGATGAAAGAAATTTCAAGTTTATGTAATATAAACAAC 1080
DB 180331 TCACGATATAGTAGACAGATGAAAGAAATTTCAAGTTTATGTAATATAAACAAC 180272
QY 1081 AAGGCTTGAATATGCTATATAGGCGGTCAAGAAAGCTTCATTGAAAGATAGCATTT 1140
DB 180271 AAGGCTTGAATATGCTATATAGGCGGTCAAGAAAGCTTCATTGAAAGATAGCATTT 180212
QY 1141 AAGCAGAGTCAAGCTAGAAATATGTAATTTCAAGTTCAAGTTTATGTTCTGGGTT 1200
DB 180211 AAGCAGAGTCAAGCTAGAAATATGTAATTTCAAGTTTCAAGTTTATGTTCTGGGTT 180152
QY 1201 GGTAAATTAAGCTTTTCCCAAGGTGAAACTACCAAGAAAGACTAATTAAGTAG 1260
DB 180151 GGTAAATTAAGCTTTTCCCAAGGTGAAACTACCAAGAAAGACTAATTAAGTAG 180092
QY 1261 TGGTGTGCTCTCTGGAAGAGACACTCTGTTTCTGCTCATTAAGTCAACCTTTC 1320
DB 180091 TGGTGTGCTCTCTGGAAGAGACACTCTGTTTCTGCTCATTAAGTCAACCTTTC 180032
QY 1321 ACTTCAGGCACTTTTGAAGCCCTTTGCCAGTCAAGGGAAGGGAGGCTGGGCATG 1380
DB 180031 ACTTCAGGCACTTTTGAAGCCCTTTGCCAGTCAAGGGAAGGGAGGCTGGGCATG 179972
QY 1381 GGGCTTGACATTTTGAACAAGTGAACATTAATTTGCCAGACTCACTAGCCCAAGGCT 1440
DB 179971 GGGCTTGACATTTTGAACAAGTGAACATTAATTTGCCAGACTCACTAGCCCAAGGCT 179912
QY 1441 AAGCTGAAGAGGCTTTGGCATGCCCCAGAAAGGCCCTGATGAAGCTTGAAGAAAGCTG 1500
DB 179911 AAGCTGAAGAGGCTTTGGCATGCCCCAGAAAGGCCCTGATGAAGCTTGAAGAAAGCTG 179852

```



REFERENCE  
AUTHORS  
2 (bases 1 to 191656)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,  
Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A.,  
Klein, J., Larocque, K., Lamazeres, R., Landers, T., Lehoczy, J.,  
Levine, R., Liu, G., Locke, K., Macdonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,  
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
Submitted (08-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS  
3 (bases 1 to 191656)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Kartas, A., Kells, C., Larocque, K., Lamazeres, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Margulis, N., Mathews, G.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,  
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Pihunghang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
Strasas, N., Subramanian, A., Talamas, J., Tessfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS  
4 (bases 1 to 191656)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Kartas, A., Kells, C., Larocque, K., Lamazeres, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Margulis, N., Mathews, G.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,  
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Pihunghang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
Strasas, N., Subramanian, A., Talamas, J., Tessfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2002 this sequence version replaced g1:19683105.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

FEATURES  
source  
-----  
Location/Qualifiers  
1. 191656  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="RP11-304C12"  
/clone\_11b="RPCT-11 Human Male BAC"  
complement(189..1091)  
/rpt\_family="L2"  
complement(1151..1401)  
/rpt\_family="L2"  
complement(1891..2187)  
/rpt\_family="L2"  
complement(3141..3165)  
/rpt\_family="L2"  
complement(3205..3522)  
/rpt\_family="L2"  
complement(4393..4591)  
/rpt\_family="L2"  
complement(4624..4932)  
/rpt\_family="L2"  
complement(4987..5077)  
/rpt\_family="L2"  
complement(5165..6245)  
/rpt\_family="L2"  
complement(6246..6555)  
/rpt\_family="L2"  
complement(6556..6634)  
/rpt\_family="L2"  
complement(6940..7136)  
/rpt\_family="L2"  
complement(7607..7849)  
/rpt\_family="L2"  
complement(11242..11318)  
/rpt\_family="L2"  
complement(11385..11572)  
/rpt\_family="L2"  
complement(12325..12520)  
/rpt\_family="L2"  
complement(12536..12706)  
/rpt\_family="L2"  
complement(12905..13529)  
/rpt\_family="L2"  
complement(13530..13842)  
/rpt\_family="L2"  
complement(13843..16076)  
/rpt\_family="L2"  
complement(14517..14528)  
/rpt\_family="L2"  
complement(14536..14542)  
/rpt\_family="L2"  
complement(14553..14558)  
/rpt\_family="L2"

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 06:39:26 ; Search time 23513.6 Seconds

(without alignments)  
17398.312 Million cell updates/secTitle: US-09-900-448-3\_COPY\_1\_10000  
Perfect score: 10000  
Sequence: 1 tccctctcccccagcagccgcccc.....agctgatttcgagcctaaac 10000Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_man:\*  
37: em\_htg\_vrc:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	9983.2	99.8	191656	9 AC068733	AC068733 Homo sapi
C 2	9973.2	99.7	112285	9 AC068433	AC068433 Homo sapi
C 3	7556.4	75.6	148198	2 AC024232	AC024232 Homo sapi
C 4	2374.8	23.7	148198	2 AC024232	AC024232 Homo sapi
C 5	1509.4	15.1	5234	6 AX345846	AX345846 Sequence
C 6	1281.6	12.8	5234	6 AX345847	AX345847 Sequence
C 7	681.4	6.8	731	9 HSMXMA01	M36796 Human hemop
C 8	509.4	5.1	511	9 HSMXMA01	M36796 Human hemop
C 9	479.6	4.8	249769	2 AC152227	X15399 Human DNA f
C 10	479.4	4.8	202495	9 CNS01DM6	AC125227 Mus muscu
C 11	471.6	4.7	74904	9 AC108045	AL136418 Human chr
C 12	467.4	4.7	202496	9 CNS01DM6	AL139054 Human chr
C 13	444.8	4.4	187360	9 AC021105	AC021105 Human chr
C 14	439.6	4.4	176287	9 HSBAL2M9	AL080243 Human DNA
C 15	439.4	4.4	8852	10 WPMPEX1	X56829 M.musculus
C 16	436.2	4.4	185909	9 AL590434	AL590434 Human DNA
C 17	430.2	4.3	248281	9 AC008737	AC008737 Homo sapi
C 18	429.	4.3	275982	2 AC131626	AC131626 Rattus no
C 19	428.	4.3	175827	9 AL358472	AL358472 Human DNA
C 20	422.4	4.2	182535	9 AC079331	AC079331 Homo sapi
C 21	422.4	4.2	216789	9 AC006534	AC006534 Homo sapi
C 22	419.4	4.2	457	9 HUMXMA02	M36798 Human hemop
C 23	415.8	4.2	148525	2 AC016675	AC016675 Homo sapi
C 24	414.6	4.1	178105	9 AC073057	AC073057 Homo sapi
C 25	414.2	4.1	57722	2 HSCB33B10	282175 Homo sapien
C 26	410.2	4.1	106008	9 HSE06	AL020993 Human DNA
C 27	408.6	4.1	160822	2 AC020920	AC020920 Homo sapi
C 28	408.4	4.1	193148	9 AP003732	AP003732 Homo sapi
C 29	405.8	4.1	36095	9 AC013256	AC013256 Homo sapi
C 30	404.4	4.0	190570	2 AL157894	AL157894 Homo sapi
C 31	403.6	4.0	185559	2 AC019166	AC019166 Homo sapi
C 32	403.4	4.0	102717	9 AC004079	AC004079 Homo sapi
C 33	403.4	4.0	164026	2 AC092762	AC092762 Pan trogl
C 34	403.	4.0	286758	9 AC006449	AC006449 Homo sapi
C 35	401.6	4.0	187383	2 AC138954	AC138954 Homo sapi
C 36	400.6	4.0	73431	2 AP000575	AP000575 Homo sapi
C 37	400.4	4.0	155824	2 AC087296	AC087296 Homo sapi
C 38	399.6	4.0	191440	9 AC012645	AC012645 Homo sapi
C 39	399.	4.0	61993	9 AP000828	AP000828 Homo sapi
C 40	398.2	4.0	146741	9 AC011452	AC011452 Homo sapi
C 41	398.	4.0	195880	9 AC074117	AC074117 Homo sapi
C 42	397.4	4.0	178459	2 AC079931	AC079931 Homo sapi
C 43	397	4.0	116478	9 AC092287	AC092287 Homo sapi
C 44	397	4.0	172876	9 AC114489	AC114489 Homo sapi
C 45	397	4.0	179221	9 AC114493	AC114493 Homo sapi

## ALIGNMENTS

RESULT 1  
AC068733/c 191656 bp DNA linear PRI 24-MAY-2002  
LOCUS AC068733 Homo sapiens chromosome 11, clone RP11-304C12, complete sequence.  
DEFINITION AC068733  
ACCESSION AC068733  
VERSION AC068733.12 GI:19703148  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.  
TITLE 1 (bases 1 to 191656)  
JOURNAL Homo sapiens chromosome 11, clone RP11-304C12  
Unpublished



**THIS PAGE BLANK (USPTO)**

CURRENT FILING DATE: 2002-04-26  
 PRIOR APPLICATION NUMBER: US 60/286,764  
 PRIOR FILING DATE: 2001-04-26  
 NUMBER OF SEQ ID NOS: 80  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 5  
 LENGTH: 606  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-133-797-5

Query Match 9.1%; Score 199; DB 15; Length 606;  
 Becc Local Similarity 25.0%; Pred. No. 3.2e-10;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY	5	IGAPVALGL-----WSLCMSLAIAATPL-----PPTSAHGUNAEGETKPDPTVTERCS	51
Db	277	VGDFLRIGLPIEDKVRWQLYGVRESVPTAQPEEPFLPEPPDNRSSAPPRKDVPHKCS	336
QY	52	DGWSFDATTTLDNGTMLFFKGEFWKSHKMDRELISER-----WKNFP---SPVDAAF	101
Db	337	T--HFDAVA-QIRGEAFFKGYFWRLLR-DRHLVSLQPAQMRFRGLPLHDSVDVAVY	392
QY	102	-RQHNSVFLIKDKVMYTPPEKKEKYPKLLQDEFPGIPSPLDAAVECHRGECQAEGL	160
Db	393	ERTSDHKIVFFKGDGRYWFKDNVNBEGYPRPVSD-FSLPPGIDAA-----	437
QY	161	FFQGHGRNGTGHGNSSTHGPYWRCSPHLVLSALTSNMGATYAFSGTHYWRIDT--SRD	219
Db	438	FSMAHNR-----TYFFKQQLYWRDYDHTRH	463
QY	220	GMHSMPI-AHQWPGSPSAVDAAFSMEB-KLYIVOGTQVYVFLTKGGYTLVSGYPRLEKE	277
Db	464	MDPGYPAQSPIMRGVPSTLDDAMRMWDGASVFFRG-QEYWKVLDSGELEVAAGYPQSTARD	522
QY	278	-----VGTPHGIIIDSVDAA	292
Db	523	WLVCGDSQADGSVAAGVDAA	542

Search completed: December 16, 2003, 06:59:58  
 Job time : 253 secs

Db 250 T--HFDVA-QIRGEAFFKGYFWRLTR-DRHLVSLQPAQMRFWRGRLPHLDSVDAY 305  
QY 102 -RQSHSVFLIKGDKVWVPEPEKKEGYPKLLODEFFGIPSLDAVVECHRGECQAEGVL 160  
Db 306 ERTSDHKIYFFKGDYRWVFKDNNVEGYPVPVD-FSLPPGIDAA----- 350  
QY 161 FFOGHGRNCTGHNSTHGGPEYMRCSPHLVLSALTSDNHATYAFSGTHYWRDLT-SRD 219  
Db 351 FSWAHNDR-----TYFFKQOLYMRDHTRH 376  
QY 220 GWHSMPI-AHQWPGPSAVDAAFSWEB-KLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKE 277  
Db 377 MDGYPAQSPLMKGVSTLDDARMSDGSAYFFRG-QEYWKVLDGELEVAAPGYPOSTARD 435  
QY 278 -----VGTPIGIIILDSVDA 292  
Db 436 WLVCSDSQADGSVAAGVDA 455

## RESULT 13

US-10-131-985-49  
; Sequence 49, Application US/10131985  
; Publication No. US2003019440A1  
; GENERAL INFORMATION:  
; APPLICANT: Dack, Kevin N  
; APPLICANT: Davies, Michael J  
; APPLICANT: Fish, Paul V  
; APPLICANT: Hugins, Jonathan P  
; APPLICANT: McIntosh, Fraser S  
; APPLICANT: Ocleston, Nicholas L  
; TITLE OF INVENTION: Composition  
; FILE REFERENCE: PCS 10391A  
; CURRENT APPLICATION NUMBER: US/10/131,985  
; CURRENT FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: US/09/726,295  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: GB 9930768.8  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 519  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-985-49

Query Match 9.1%; Score 199; DB 12; Length 519;  
Best Local Similarity 25.0%; Pred. No. 2.6e-10;

Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----WSLCWSLAIATPL-----PPTSAHGNVABGETKPPDDYTERCS 51  
Db 190 VGOPLRYGLYEDKVRWQLYGVRESVPTAQBEPPEPLPEPPDNRSSAPPRKDVHRC 249  
QY 52 DGWSPDATTLDNGTMLFFKGEFVWKSMDRELISER-----WKNRP---SPVDAAF 101  
Db 250 T--HFDVA-QIRGEAFFKGYFWRLTR-DRHLVSLQPAQMRFWRGRLPHLDSVDAY 305  
QY 102 -RQSHSVFLIKGDKVWVPEPEKKEGYPKLLODEFFGIPSLDAVVECHRGECQAEGVL 160  
Db 306 ERTSDHKIYFFKGDYRWVFKDNNVEGYPVPVD-FSLPPGIDAA----- 350  
QY 161 FFOGHGRNCTGHNSTHGGPEYMRCSPHLVLSALTSDNHATYAFSGTHYWRDLT-SRD 219  
Db 351 FSWAHNDR-----TYFFKQOLYMRDHTRH 376  
QY 220 GWHSMPI-AHQWPGPSAVDAAFSWEB-KLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKE 277  
Db 377 MDGYPAQSPLMKGVSTLDDARMSDGSAYFFRG-QEYWKVLDGELEVAAPGYPOSTARD 435  
QY 278 -----VGTPIGIIILDSVDA 292  
Db 436 WLVCSDSQADGSVAAGVDA 455

## RESULT 14

US-10-406-209-2  
; Sequence 2, Application US/10406209  
; Publication No. US20030170758A1  
; GENERAL INFORMATION:  
; APPLICANT: KYOMA HAKKO KOGYO CO., LTD.  
; TITLE OF INVENTION: NOVEL ANTIBODIES, DRUGS CONTAINING THESE ANTIBODIES AND  
; TITLE OF INVENTION: METHODS FOR  
; FILE REFERENCE: 1241.19  
; CURRENT APPLICATION NUMBER: US/10/406,209  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: US/09/806,228C  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/JP99/05350  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291501  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291503  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-406-209-2

Query Match 9.1%; Score 199; DB 12; Length 605;  
Best Local Similarity 25.0%; Pred. No. 3.2e-10;

Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----WSLCWSLAIATPL-----PPTSAHGNVABGETKPPDDYTERCS 51  
Db 276 VGOPLRYGLYEDKVRWQLYGVRESVPTAQBEPPEPLPEPPDNRSSAPPRKDVHRC 335  
QY 52 DGWSPDATTLDNGTMLFFKGEFVWKSMDRELISER-----WKNRP---SPVDAAF 101  
Db 336 T--HFDVA-QIRGEAFFKGYFWRLTR-DRHLVSLQPAQMRFWRGRLPHLDSVDAY 391  
QY 102 -RQSHSVFLIKGDKVWVPEPEKKEGYPKLLODEFFGIPSLDAVVECHRGECQAEGVL 160  
Db 392 ERTSDHKIYFFKGDYRWVFKDNNVEGYPVPVD-FSLPPGIDAA----- 436  
QY 161 FFOGHGRNCTGHNSTHGGPEYMRCSPHLVLSALTSDNHATYAFSGTHYWRDLT-SRD 219  
Db 437 FSWAHNDR-----TYFFKQOLYMRDHTRH 462  
QY 220 GWHSMPI-AHQWPGPSAVDAAFSWEB-KLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKE 277  
Db 463 MDGYPAQSPLMKGVSTLDDARMSDGSAYFFRG-QEYWKVLDGELEVAAPGYPOSTARD 521  
QY 278 -----VGTPIGIIILDSVDA 292  
Db 522 WLVCSDSQADGSVAAGVDA 541

## RESULT 15

US-10-133-797-5  
; Sequence 5, Application US/10133797  
; Publication No. US20030109021A1  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Shujian  
; APPLICANT: Chen, Jian  
; APPLICANT: Feder, John  
; APPLICANT: Lee, Liana  
; APPLICANT: Krystek, Stanley  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY  
; FILE REFERENCE: D0141NP  
; CURRENT APPLICATION NUMBER: US/10/133,797

QY 202 ATYAFSGTHYRLDTSRD---GHSWPIAHQWPGPSAVDAAFSWEK-LYLVQGTQVY 256  
Db 392 -TYFVDNQWRYRDERBQMDPGYFKLITKNFGGIGPK-IDAVFYSKNYYFFQGSNOF 449  
QY 257 VF 258  
Db 450 EY 451

RESULT 10

US-10-171-311-135  
; Sequence 135, Application US/10171311  
; Publication No. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatc, Karen  
; APPLICANT: Gamaavarapu, Manjula  
; APPLICANT: Hoersch, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; TITLE OF INVENTION: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035  
; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 135  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-135

Query Match 9.3%; Score 202.5; DB 15; Length 470;  
Best Local Similarity 27.3%; Pred. No. 1.1e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGVAEGETKPPDVTY--RCSGWSFDATTLDDNGTMLFKGGEFVW---KSHKMDR 83  
Db 260 SLYDPPKRNQRLPNPDNSEPALCDPNLSFDVAVTVGN-KIFFFKDRFFWLKVSERPKTSV 318  
QY 84 ELISERWKNFSPVPDAFR-OGHNSVFLIKGDKVWVYPPKEKKGYPKLLD-BPPGIPS 141  
Db 319 NLISSLWTLPSGIEAIEIARNOVFLPKDKTWLISNLAPENPYPSISFGFENPVK 378  
QY 142 PLDAVBECHRGCEQAEGLVLFQGHGHRNGTGHGNSHTHGPEYMCSPHLVLSALTSNMG 201  
Db 379 KIDAAV-----FNPRFYR----- 391  
QY 202 ATYAFSGTHYRLDTSRD---GHSWPIAHQWPGPSAVDAAFSWEK-LYLVQGTQVY 256  
Db 392 -TYFVDNQWRYRDERBQMDPGYFKLITKNFGGIGPK-IDAVFYSKNYYFFQGSNOF 449  
QY 257 VF 258  
Db 450 EY 451

RESULT 11  
US-09-925-301-1217  
; Sequence 1217, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1217  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1217

Query Match 9.3%; Score 202.5; DB 9; Length 473;  
Best Local Similarity 27.3%; Pred. No. 1.1e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGVAEGETKPPDVTY--RCSGWSFDATTLDDNGTMLFKGGEFVW---KSHKMDR 83  
Db 263 SLYDPPKRNQRLPNPDNSEPALCDPNLSFDVAVTVGN-KIFFFKDRFFWLKVSERPKTSV 321  
QY 84 ELISERWKNFSPVPDAFR-OGHNSVFLIKGDKVWVYPPKEKKGYPKLLD-BPPGIPS 141  
Db 322 NLISSLWTLPSGIEAIEIARNOVFLPKDKTWLISNLAPENPYPSISFGFENPVK 381  
QY 142 PLDAVBECHRGCEQAEGLVLFQGHGHRNGTGHGNSHTHGPEYMCSPHLVLSALTSNMG 201  
Db 382 KIDAAV-----FNPRFYR----- 394  
QY 202 ATYAFSGTHYRLDTSRD---GHSWPIAHQWPGPSAVDAAFSWEK-LYLVQGTQVY 256  
Db 395 -TYFVDNQWRYRDERBQMDPGYFKLITKNFGGIGPK-IDAVFYSKNYYFFQGSNOF 452  
QY 257 VF 258  
Db 453 EY 454

RESULT 12  
US-09-801-196-30  
; Sequence 30, Application US/09801196  
; Patent No. US20020037827A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Kai  
; APPLICANT: Smith, Ryan  
; APPLICANT: Fajardo, Mark  
; APPLICANT: Moss, Patrick  
; TITLE OF INVENTION: A NOVEL MATRIX METALLOPROTEINASE (MMP-25)  
; TITLE OF INVENTION: EXPRESSED IN SKIN CELLS  
; FILE REFERENCE: 240083.509  
; CURRENT APPLICATION NUMBER: US/09/801,196  
; CURRENT FILING DATE: 2001-03-06  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 519  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-801-196-30

Query Match 9.1%; Score 199; DB 9; Length 519;  
Best Local Similarity 25.0%; Pred. No. 2.6e-10;  
Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----MSLCWSLAIATPL-----PPTSAGHVAEGETKPPDVTERRCS 51  
Db 190 VGDPURGLPYEDKVRWQLYGVRESVPTAQPEPPLLEBPDPNRSAPPRKDVPRCS 249  
QY 52 DGWSFDATTLDDNGTMLFKGGEFVWKSMDRELISER-----WKNFP---SPVDAAF 101



PRIOR FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/JP99/05350  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291501  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-29  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291503  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-29  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 587  
TYPE: PRT  
ORGANISM: Mouse  
US-10-406-209-1

Query Match 9.3%; Score 204; DB 12; Length 587;  
Best Local Similarity 24.4%; Pred. No. 1e-10; Indels 158; Gaps 25;  
Matches 103; Conservative 34; Mismatches 127;

QY 5 LGATVALGL-----WSLCMSLAATPL-----PPTSANGVAGETKPDV 46  
DB 273 VGDVRYGLPYEDRYVWQLYGVRESVPTAQLOTPEPEEPFLPEPPNNSSTPPQKDV 332  
QY 47 TERCSGKSPDATTLDDNGTMLPFKGEFVWKS HKMDRELISER-----WKNP--SP 96  
DB 333 PHRCTA--HFDVAVA-QIRGEAFFEKGYFMRLTR--DRHLVSLQPAQMERFWRGLPLHDS 388  
QY 97 VDAAF-RQGHNSVFLIKGDKVWVYPREKKEGYPLLODEFPGLSPILDAVECHRGSCQ 155  
DB 389 VDAVYERTSDKITYFPGKDRYVWFQDNNEEGYPRPSVD--PSLPGGIDA-----437  
QY 156 AEGVLFFQGHGRNGTGHGNSHTHGPBYMRCSPHLVLSALTSDNGATVAFSGTHYMRLD 215  
DB 438 ----VFSANHDR-----TYFFKQDQLYMYXD 459  
QY 216 --TSR--DQWHS-MPIAHQWPGQSAVDAAPSWEB-KLYLVQGTQVYVFLTKGTYTVSG 269  
DB 460 DHTRMOPGYPAQGPL--WRGVPSMLDAMRWSDGASYPFRG-QEYWKVLDDGELEAARG 515  
QY 270 YPKRLEKE--VGTPHGIIIDSVDAAFICPGSSRHIMAGRIMWLDLKSQAQATWETLP 326  
DB 516 YPOSTARDWLVGEP-----LADAEVGPBPQ-----GR-----SQAQ-----548  
QY 327 WPHEKVDG-ALCMESLGPNSCSANGPGLYLIHGPNLYCYSDVEKLANAKALPOQVNTS 385  
DB 549 -----DGLAVCS-----CTSDARL-----ALPSILLITP 573  
QY 386 LL 387  
DB 574 LL 575

RESULT 5  
US-09-801-196-23  
Sequence 23, Application US/09801196  
Patent No. US20020037827A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Kai  
APPLICANT: Smith, Ryan  
APPLICANT: Fajardo, Mark  
APPLICANT: Moss, Patrick  
TITLE OF INVENTION: A NOVEL MATRIX METALLOPROTEINASE (MMP-25)  
FILE REFERENCE: 240083.509  
CURRENT APPLICATION NUMBER: US/09/801.196  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 470  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-801-196-23

Query Match 9.3%; Score 202.5; DB 9; Length 470;  
Best Local Similarity 27.3%; Pred. No. 1.1e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAGETKPPDYTE--RCSDGWSFDAATLDDNGTMLFFKGEFVW----KSHKMDR 83  
DB 260 SLYGDPKRNQRLPNPDSBPALCPNLSFDVAVTVGN-KIFFKDRFFWLKVSRPKTSV 318  
QY 84 ELISERWKNFSPVDAAFR-QGHNSVFLIKGDKVWVYPREKKEGYPLLOD-EFPGLPS 141  
DB 319 NLISSLMPTLPDSGLEAAVEIARNQVFLPKDQKWLISNLRPEBNYPSKSHSFPQPNVK 378  
QY 142 PLDAVECHRGCEQAEGLVFFQGHGRNGTGHGNSHTHGPBYMRCSPHLVLSALTSDNHG 201  
DB 379 KIDAAN-----FNPRFYR-----391  
QY 202 ATYAFSGTHYWRDLTSD--GHSWPIAHQWPGQSAVDAAPSWEB-KLYLVQGTQVY 256  
DB 392 -TYFFVNDQYWRDYERRQMDPGYPKLITKNFGQIGPK-IDAVFYSKKYYFFQGSNOF 449  
QY 257 VF 258  
DB 450 EY 451

RESULT 6  
US-09-920-455-219  
Sequence 219, Application US/09920455  
Patent No. US20020168647A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
APPLICANT: Fan, Liqun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.540  
CURRENT APPLICATION NUMBER: US/09/920.455  
CURRENT FILING DATE: 2001-08-01  
NUMBER OF SEQ ID NOS: 275  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 219  
LENGTH: 470  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-920-455-219

Query Match 9.3%; Score 202.5; DB 10; Length 470;  
Best Local Similarity 27.3%; Pred. No. 1.1e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAGETKPPDYTE--RCSDGWSFDAATLDDNGTMLFFKGEFVW----KSHKMDR 83  
DB 260 SLYGDPKRNQRLPNPDSBPALCPNLSFDVAVTVGN-KIFFKDRFFWLKVSRPKTSV 318  
QY 84 ELISERWKNFSPVDAAFR-QGHNSVFLIKGDKVWVYPREKKEGYPLLOD-EFPGLPS 141  
DB 319 NLISSLMPTLPDSGLEAAVEIARNQVFLPKDQKWLISNLRPEBNYPSKSHSFPQPNVK 378  
QY 142 PLDAVECHRGCEQAEGLVFFQGHGRNGTGHGNSHTHGPBYMRCSPHLVLSALTSDNHG 201  
DB 379 KIDAAN-----FNPRFYR-----391  
QY 202 ATYAFSGTHYWRDLTSD--GHSWPIAHQWPGQSAVDAAPSWEB-KLYLVQGTQVY 256  
DB 392 -TYFFVNDQYWRDYERRQMDPGYPKLITKNFGQIGPK-IDAVFYSKKYYFFQGSNOF 449  
QY 257 VF 258  
DB 450 EY 451

RESULT 7  
US-10-301-822-123

Db 181 PEYMRCSPLVLVSALTSDNHGATVAFSGTHYRLDTSRDCMSWPIAHQWPGPSAVDA 240  
 Qy 241 FSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIILDSVDAATICGSSR 300  
 Db 241 FSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIILDSVDAATICGSSR 300  
 Qy 301 LHMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNSCSANGPGLYLHGP 360  
 Db 301 LHMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNSCSANGPGLYLHGP 360  
 Qy 361 NLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 391  
 Db 361 NLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 391

RESULT 2  
 US-09-900-448-4  
 ; Sequence 4, Application US/09900448  
 ; Publication No. US20030220488A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CECCARDI, Toni et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
 ; FILE OF INVENTION: US/09/900,448  
 ; FILE REFERENCE: CL001272  
 ; CURRENT FILING DATE: 2001-07-09  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 462  
 ; TYPE: PRP  
 ; ORGANISM: Human  
 ; US-09-900-448-4

Query Match 97.9%; Score 2139.5; DB 12; Length 462;  
 Best Local Similarity 84.6%; Pred. No. 1.2e-196;  
 Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;

Qy 1 MAVVLGAPVALGLMSLCMSLAITPLPPTSAGNVAGETKPPDVTYERCSDGMSFPATT 60  
 Db 1 MAVVLGAPVALGLMSLCMSLAITPLPPTSAGNVAGETKPPDVTYERCSDGMSFPATT 60  
 Qy 61 LDNGMTLFPKGFVWMSHKMDRELISERWKNFSPVDAAPROGNSVFLIKGDKVWVP 120  
 Db 61 LDNGMTLFPKGFVWMSHKMDRELISERWKNFSPVDAAPROGNSVFLIKGDKVWVP 120  
 Qy 121 PEKKEKGYPKLLQDEFFGIPSPDLDAVECHRGECQAEGLVFFQ----- 163  
 Db 121 PEKKEKGYPKLLQDEFFGIPSPDLDAVECHRGECQAEGLVFFQ----- 163  
 Qy 164 -----GHGHRN 169  
 Db 181 SWPAVNGCSALRWLGRRYCFQGNQFLRPDPVGEVPPRYRDPVDFPMPGPGHGRN 240  
 Qy 170 GTGHGNSHTHGPPEYMRCSPLVLVSALTSDNHGATVAFSGTHYRLDTSRDCMSWPIAHQ 229  
 Db 241 GTGHGNSHTHGPPEYMRCSPLVLVSALTSDNHGATVAFSGTHYRLDTSRDCMSWPIAHQ 300  
 Qy 230 WPGPSAVDAAFSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIILDSV 289  
 Db 301 WPGPSAVDAAFSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIILDSV 360  
 Qy 290 DAAPICGSSRLHIMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNSCSA 349  
 Db 361 DAAPICGSSRLHIMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNSCSA 420  
 Qy 350 NGPGLYLHGPNLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 391  
 Db 421 NGPGLYLHGPNLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 462

RESULT 3

US-10-316-253-40  
 ; Sequence 40, Application US/10316253  
 ; Publication No. US20030162706A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Procter & Gamble Company  
 ; APPLICANT: Peters, Kevin  
 ; APPLICANT: Thompson, Larry  
 ; APPLICANT: Wang, Peng  
 ; APPLICANT: Greis, Kenneth  
 ; TITLE OF INVENTION: Angiogenesis Modulating Proteins  
 ; FILE REFERENCE: 8865M  
 ; CURRENT APPLICATION NUMBER: US/10/316,253  
 ; CURRENT FILING DATE: 2002-12-10  
 ; PRIOR APPLICATION NUMBER: US 60/355,295  
 ; NUMBER OF SEQ ID NOS: 308  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 40  
 ; LENGTH: 460  
 ; TYPE: PRP  
 ; ORGANISM: Rattus norvegicus  
 ; US-10-316-253-40

Query Match 71.2%; Score 1556.5; DB 12; Length 460;  
 Best Local Similarity 63.8%; Pred. No. 1.1e-140;  
 Matches 296; Conservative 35; Mismatches 54; Indels 79; Gaps 6;

Qy 1 MAVVLGAPVALGLMSLCMSLAITPLPPTSAGNVAGETKPPDVTYERCSDGMSFPDA 58  
 Db 1 MAVVLGAPVALGLMSLCMSLAITPLPPTSAGNVAGETKPPDVTYERCSDGMSFPDA 58  
 Qy 59 TTLDNGMTLFPKGFVWMSHKMDRELISERWKNFSPVDAAPROGNSVFLIKGDKVWVP 118  
 Db 59 TTLDNGMTLFPKGFVWMSHKMDRELISERWKNFSPVDAAPROGNSVFLIKGDKVWVP 117  
 Qy 119 YPEKKEKGYPKLLQDEFFGIPSPDLDAVECHRGECQAEGLVFFQCH----- 165  
 Db 118 YPEKKEKGYPKLLQDEFFGIPSPDLDAVECHRGECQAEGLVFFQCH----- 177  
 Qy 166 -----GH-- 167  
 Db 178 ERSWPAVNGCTALRWLGRRYCFQGNKFLRPDPVGEVPPRYRDPVDFPMPGPGHGRK 237  
 Qy 168 -RNGTHGNSHTHGPPEYMRCSPLVLVSALTSDNHGATVAFSGTHYRLDTSRDCMSWPI 226  
 Db 238 -RNGTHGNSHTHGPPEYMRCSPLVLVSALTSDNHGATVAFSGTHYRLDTSRDCMSWPI 225  
 Qy 227 AHQWPGPSAVDAAFSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIIL 286  
 Db 226 AHQWPGPSAVDAAFSWEEKLYLVQGVVFLTKGGYTLVSGYPRKLEKEVGTGPHIIL 355  
 Qy 287 DSVDAAPICGSSRLHIMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNS 346  
 Db 356 DSVDAAPICGSSRLHIMAGRRLMWLDLKSQAQATWTELPMPHEKVDGALCMESLGPNS 415  
 Qy 347 CSANGPGLYLHGPNLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 390  
 Db 416 CSANGPGLYLHGPNLVYCSYDVEKLNAAKALPOPOVNTSLGCTH 459

RESULT 4  
 US-10-406-209-1  
 ; Sequence 1, Application US/10406209  
 ; Publication No. US20030170758A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KYOMA HAKKO KOGYO CO., LTD.  
 ; TITLE OF INVENTION: NOVEL ANTIBODIES, DRUGS CONTAINING THESE ANTIBODIES AND  
 ; TITLE OF INVENTION: METHODS FOR  
 ; FILE REFERENCE: 1241.19  
 ; CURRENT APPLICATION NUMBER: US/10/406,209  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: US/09/806,228C



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 06:48:58 ; Search time 252 Seconds  
(without alignments)  
288.570 Million cell updates/sec

Title: US-09-900-448-2  
Perfect score: 2185  
Sequence: 1 MARVLGAPVALGLMSLCWSL.....NAKALPQPNVTSLLGCTH 391

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues  
Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2185	100.0	391	12	US-09-900-448-2
2	2139.5	97.9	462	12	US-09-900-448-4
3	1556.5	71.2	460	12	US-10-316-253-40
4	204	9.3	587	12	US-10-406-209-1
5	202.5	9.3	470	9	US-09-801-196-23
6	202.5	9.3	470	10	US-09-920-455-219
7	202.5	9.3	470	12	US-10-301-822-123
8	202.5	9.3	470	12	US-10-337-632-2
9	202.5	9.3	470	12	US-10-131-985-39
10	202.5	9.3	470	15	US-10-171-311-135
11	202.5	9.3	473	9	US-09-925-301-1217
12	199	9.1	519	9	US-09-801-196-30
13	199	9.1	519	12	US-10-131-985-49
14	199	9.1	605	12	US-10-406-209-2
15	199	9.1	606	15	US-10-133-797-5

16	195.5	8.9	477	10	US-09-920-455-260	Sequence 260, App
17	189.5	8.7	476	9	US-09-801-196-25	Sequence 25, App1
18	189.5	8.7	476	12	US-10-021-660-72	Sequence 72, App1
19	189.5	8.7	476	12	US-10-131-985-35	Sequence 35, App1
20	183.5	8.4	467	9	US-09-801-196-20	Sequence 20, App
21	183.5	8.4	467	11	US-09-759-1308-176	Sequence 176, App
22	183.5	8.4	467	12	US-10-131-985-31	Sequence 31, App1
23	181.5	8.3	469	9	US-09-801-196-19	Sequence 19, App1
24	181.5	8.3	469	9	US-09-853-386-100	Sequence 100, App
25	181.5	8.3	469	12	US-10-301-822-119	Sequence 119, App
26	181.5	8.3	469	12	US-10-021-660-76	Sequence 76, App1
27	181.5	8.3	469	12	US-10-308-279-34	Sequence 34, App1
28	181.5	8.3	469	12	US-10-131-985-23	Sequence 23, App1
29	181.5	8.3	496	15	US-10-106-698-6283	Sequence 6283, Ap
30	179.5	8.2	488	9	US-09-801-196-26	Sequence 26, App1
31	179.5	8.2	488	12	US-10-301-822-121	Sequence 121, App
32	179.5	8.2	488	12	US-10-131-985-37	Sequence 37, App1
33	179.5	8.2	488	15	US-10-177-293-298	Sequence 298, App
34	179.5	8.2	488	16	US-10-225-486-67	Sequence 67, App1
35	177	8.1	477	9	US-09-801-196-24	Sequence 24, App1
36	177	8.1	477	12	US-10-301-822-127	Sequence 127, App
37	177	8.1	477	12	US-10-131-985-27	Sequence 27, App1
38	177	8.1	477	15	US-10-171-311-137	Sequence 137, App
39	176	8.1	471	9	US-09-801-196-31	Sequence 31, App1
40	176	8.1	471	9	US-09-801-196-32	Sequence 32, App1
41	176	8.1	471	12	US-10-375-884-1	Sequence 1, App1
42	176	8.1	471	12	US-10-375-884-21	Sequence 21, App1
43	176	8.1	471	12	US-10-131-985-41	Sequence 41, App1
44	176	8.1	471	15	US-10-050-2168-6	Sequence 6, App1
45	176	8.1	471	15	US-10-075-069-1	Sequence 1, App1

ALIGNMENTS

RESULT 1  
US-09-900-448-2  
; Sequence 2, Application US/09900448  
; Publication No. US20030220488A1  
GENERAL INFORMATION:  
; APPLICANT: CECARDI, Toni et al.  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; FILE REFERENCE: CL001272  
; CURRENT APPLICATION NUMBER: US/09/900,448  
; CURRENT FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Human  
US-09-900-448-2

Query Match	100.0%	Score 2185;	DB 12;	Length 391;
Best Local Similarity	100.0%	Pred. No. 4e-201;		
Matches 391;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MARVLGAPVALGLMSLCWSLIAIATPLPTSAHGNVAEGETKPPDPVTERCSGMSFDATT	60	
DB	1	MARVLGAPVALGLMSLCWSLIAIATPLPTSAHGNVAEGETKPPDPVTERCSGMSFDATT	60	
QY	61	LDNGMTLFPKSGFVWKSMDRELISERKKNPPSPVDAAFRGHNSVFLIKDKVWVP	120	
DB	61	LDNGMTLFPKSGFVWKSMDRELISERKKNPPSPVDAAFRGHNSVFLIKDKVWVP	120	
QY	121	PEKKEKYPKLLDDEFGIPSPDAAVECHRGSCQAGVFPQGHGRNGTGHGNSTHHG	180	
DB	121	PEKKEKYPKLLDDEFGIPSPDAAVECHRGSCQAGVFPQGHGRNGTGHGNSTHHG	180	
QY	181	PEYWRCSPHLVLSALTSNDHGAATYAFSGTHYWRLDTSRDGMSWPIAHQWPGSPSAVDA	240	
DB	181	PEYWRCSPHLVLSALTSNDHGAATYAFSGTHYWRLDTSRDGMSWPIAHQWPGSPSAVDA	240	

**THIS PAGE BLANK (USPTO)**

```

Db      266 TPSTPTLSSQAGTDTNEIALQEEVEPEVCET--SPDAVS--TIRGELFFFKAGFWRL 322
Qy      79 HKMDRE-----LISERWKNFPSPVDAAFROGHNSVFLIKDKVWVYPEPEKKEK 133
Db      323 RSGQLQPGVYRPLASRHWQGLPSPVDAAFEDAQGIWFGQAQYVWYDEKEKPLGPAFLSK 382
Qy      134 DEFGIPSPDLDAVECHRGECQAEGLVFFQGHGHRNGTGHGNSTHGPEYWRCSPHLVLS 193
Db      383 LGLOG--SPVHAALV-----WGPE----- 399
Qy      194 ALTSNHNATYAFSGTHYWRDL--TSR--DGMHSPHIAQ--WPGGSAVDAAFSWEKL 247
Db      400 -----KKKIYFFRGDYMRFHPTQRYVDN--PVPRRTDWRGVPSEIDAAF----- 443
Qy      248 YLVQGTQVYVFLTKG 262
Db      444 ---QDADGVAYFLRG 455

```

## RESULT 15

```

S13423
stromelysin 3 (EC 3.4.24.-) precursor - human
N/Alternate names: matrix metalloproteinase 11 (MMP11)
C/Species: Homo sapiens (man)
C/Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C/Accession: S13423; 138250; S58912
R/Basset, P.; Bellocq, J.P.; Wolf, C.; Stoll, I.; Hutin, P.; Limacher, J.M.; Podhajcer,
Nature 348, 699-704, 1990
A/Title: A novel metalloproteinase gene specifically expressed in stromal cells of breast
A/Reference number: S13423; MUID:91080920; PMID:1701851
A/Accession: S13423
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-488 <BAS>
A/Cross-references: EMBL:X57766
R/Anglard, P.; Melot, T.; Guerin, E.; Thomas, G.; Basset, P.
J. Biol. Chem. 270, 20337-20344, 1995
A/Title: Structure and promoter characterization of the human stromelysin-3 gene.
A/Reference number: 138250; MUID:95386471; PMID:7657606
A/Accession: 138250
A/Status: preliminary; translation not shown
A/Molecule type: DNA
A/Residues: 1-36 <RES>
A/Cross-references: EMBL:X84664; NID:G984746; PIDN:CAAS9150.1; PID:G987949
R/Pel, D.; Welts, S.J.
Nature 375, 244-247, 1995
A/Title: Furin-dependent intracellular activation of the human stromelysin-3 zymogen.
A/Reference number: S58912; MUID:95265105; PMID:7746327
A/Accession: S58912
A/Status: preliminary
A/Molecule type: protein
A/Residues: 81-101 <PEI>
C/Genetics:
A/Gene: GDB:MMP11; STWY3
A/Cross-references: GDB:128630; OMIM:185261
A/Map position: 22q11.2-22q11.2
C/Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprote-
C/Keywords: hydrolase; metalloproteinase; zinc; zymogen
F/48-256/Domain: matrix metalloproteinase homology <MPX>
F/288-480/Domain: hemopexin repeat homology <PXX>
F/80,215,219,225/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F/215,219,225/Binding site: zinc, catalytic (His) (active) #status predicted
F/216/Active site: Glu #status predicted

```

## Query Match

Best local similarity 8.2%; Score 179.5; DB 2; Length 488;  
Matches 65; Conservative 23; Mismatches 100; Indels 71; Gaps 12;

```

Qy      13 LMSLCWGLATATPLPPTSAHGNVAEGETK--DPDYTERCSGWSFDPATITLDNGTMLPFK 71
Db      256 LVGGPWP-TVTSRTPALGPAGIDTNEIAPLEBPADPADCA--SPDAVS--TIRGELFFFK 312

```

```

Qy      72 GEPVWKSHKMDRE-----LISERWKNFPSPVDAAFROGHNSVFLIKDKVWVYPEPEKKEK 126
Db      313 AGFWWRRLRGQLQGYRPLASRHWQGLPSPVDAAFEDAQGIWFGQAQYVWYDEKEKPL 372
Qy      127 GYPLLDDEFPGIISPLDAVECHRGECQAEGLVFFQGHGHRNGTGHGNSTHGPEYWR 186
Db      373 G-----PAPLTEL-----GLVRFVHA-----ALWGPE----- 396
Qy      187 SPHLVLSALTSNHNATYAFSGTHYWRDLTSRDGHSWPI--AHQWPGGSAVDAAFSW 243
Db      397 -----KKKIYFFRGDYMRFHPTQRYVDN--PVPRRTDWRGVPSEIDAAF-- 440
Qy      244 BEKLVQGTQVYVFLTKG 262
Db      441 ---QDADGVAYFLRG 452

```

Search completed: December 16, 2003, 06:53:08  
Job time : 38 secs

A:/Accession: S10595  
A:/Molecule type: protein  
A:/Residues: 20-107 <SUZ>  
R:/Suzuki, K.; Lee, M.; Newlands, G.F.J.; Nagase, H.; Woollley, D.E.  
Biochem. J. 305, 301-306, 1995  
A:/Title: Activation of precursors for matrix metalloproteinases 1 (interstitial collagenase)  
A:/Reference number: SS3438; PMID:59126921; PMID:7826345  
A:/Accession: SS3438  
A:/Status: preliminary  
A:/Molecule type: protein  
A:/Residues: 20-108 <SU2>  
R:/Sprigman, E.B.; Angleton, E.L.; Birkedal-Hansen, H.; Van Wart, H.E.  
Proc. Natl. Acad. Sci. U.S.A. 87, 364-368, 1990  
A:/Title: Multiple modes of activation of latent human fibroblast collagenase: evidence from A:/Reference number: A44517; PMID:90115877; PMID:2153297  
A:/Contents: annotation; disulfide bond; activation mechanism  
R:/Selow, S.P.; Marcy, A.I.; Cuca, G.C.; Smith, C.K.; Kopka, I.E.; Hagmann, W.K.; Hermes Biochemistry 31, 4535-4540, 1992  
A:/Title: Characterization of zinc-binding sites in human stromelysin-1: stoichiometry of A:/Reference number: A43031; PMID:92256384; PMID:1581308  
A:/Contents: annotation; zinc ligand in proenzyme  
A:/Note: Cys-92 binds zinc in the proenzyme. Both active and proenzyme forms of the cathepsin Comment: Procollagenase can be activated without removal of the activation peptide. St cation peptide by other proteinases.  
C:/Comment: Procollagenase is found in glycosylated and unglycosylated forms, both of which C:/Genetics:  
A:/Gene: GDB:MMP1; CLG  
A:/Cross-references: GDB:119783; OMIM:120353  
A:/Map position: 11q22.2-11q22.3  
C:/Function:  
A:/Description: hydrolyzes collagen, in particular types I, II, III, and X, serpins, and C:/Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprotein C;/Keywords: calcium; extracellular matrix; fibroblast; glycoprotein; hydrolase; metallopro F:/1-19/Domain: signal sequence #status predicted <SIG>  
F:/20-466/Product: procollagenase #status experimental <PRO>  
F:/20-99/Domain: activation peptide #status experimental <ACT>  
F:/60-261/Domain: matrix metalloproteinase homology <MMP>  
F:/90-97/Region: autoinhibitory  
F:/100-469/Product: interstitial collagenase #status experimental <MAT>  
F:/272-466/Domain: hemopexin repeat homology <PNX>  
F:/92-218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status F:/10,143/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:/218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
F:/219/Active site: Glu #status predicted  
F:/269-270/Cleavage site: Pro-Ile (autolytic) #status experimental  
F:/278-466/Disulfide bonds: #status experimental

Query Match      8.3%; Score 181.5; DB 1; Length 469;  
Best Local Similarity    26.3%; Pred. No. 2.2e+07;  
Matches    59; Conservative    26; Mismatches    82; Indels    57; Gaps    8;

Oy     PDVTERSGSDGSPPA--TLLDNGTMLFPKGFVKSHKHDEL---ISERKNPSPVD 98  
         :|||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::  
Db     PQTAKADSKITFTYHITI--RGVMEFKORFYKRTPRYEVLNITSVMPLDRGLE 329  
Oy     99 AAFA-QGHNSVFLLKGDKRWVVYPPEKEKGYPKLLODEFFPGIPAPLDAAVECHREGCOAE 157  
Db     330 AAAYEFADRBEVRPFQGNKYMAVGONLVLAGVPKDIYSF-GEP----- 371  
Oy     158 GVLFPOQHGRNRGGHGNGSTHHGEPRYCPSHLVLSALTSNHAGATYAFGSTHYWRD-- 215  
Db     372 -----RTVGH-IDALISENTGKTTFPVANKWRYDEV 403

RESULT 13  
stromelysin 3 (EC 3.4.24.-) - mouse  
A44399  
N/AIternate names: matrix metalloproteinase 11 (MMP11)  
C/Species: Mus musculus (house mouse)

C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #ext\_change 10-Jul-1998  
C/Accession: A44399  
R/Defeature: O.; Wolfe, C.; Limacher, J.M.; Hutin, P.; Wendling, C.; Leheut, M.; Basset, P.  
J. Cell Biol. 119, 997-1002, 1992  
A>Title: The breast cancer-associated stromelysin-3 gene is expressed during mouse mammary  
A/Reference number: A44399; MUID:93054930; PMID:1429845  
A/Accession: A44399  
A>Status: preliminary; not compared with conceptual translation  
A/Molecule type: nucleic acid  
A/Residues: 1-492 <LEP>  
C/Note: sequence extracted from NCBI backbone (NCBIP:117216)  
C/Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloproteinase  
C/Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F:52-262/Domain: matrix metalloproteinase homology <MMP>  
F:252-484/Domain: hemopexin repeat homology <HPN>  
F:84-219,223,229/Binding site: zinc, catalytic (His, His, His, His) (inhibited) #status predicted  
F:219,223,229/Binding site: zinc, catalytic (His) (active) #status predicted  
F:220/Active site: Glu #status predicted

Query Match 8.3%; Score 181.5; DB 2; Length 492;  
Best Local Similarity 25.7%; Pred. No. 2,4e-07;  
Matches 69; Conservative 26; Mismatches 86; Indels 93; Gaps 13;

Cy 24 TRLPPTSAH-----GNVAGETKPPDVTYRSCDSGSPATTLDDNGTMLFFKGEFVWKS 78  
Db 270 SPAPLPSQAGTDTNEILLEPETPPVCET-----SPDAVS-TTIGELTFKAGFVWRL 323  
Cy 79 HKWDR-----LISERKKNFSPVDAAFRGHSNVELIKGDKVYVPEKKEKGYPLLO 133  
Db 324 RSGRLQPEYPLALSHRWGGLPSPVDAFEDAQGIQIFGQAGYVYDGEKPYVLGPAPLSK 383  
Cy 134 DEFGIPEPLDAVNECHRGSCQAEGLVFFQGHGHRNGTGHGNSTHGPEYRCSPHVLS 193  
Db 384 LGLGQ--SPVHAALV-----WGPE----- 400  
Cy 194 ALTSDNHGATYAFSGSTHYWRLD--TSR-DGWSMPIAHQ--WPQSPAVDAAFSWEKL 247  
Db 401 -----KKKITYFRGQDYWRFRHFRQVDN---PYPRSTDKRGVPSIEDAIF----- 444  
Cy 248 YLVQSTQYVYVFLTKG-----GYTVSGYK 272  
Db 445 ---QDAEGYAFRLRHLWYMKRDPVKKVYKLEGFPR 475

RESULT 14  
JC6197  
stromelysin 3 (EC 3.4.24.-) - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #ext\_change 17-Mar-1999  
C/Accession: JC6197  
R/Okada, A.; Saez, S.; Mismul, Y.; Basset, P.  
Gene 185, 187-193, 1997  
A>Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin and  
A/Reference number: JC6197; MUID:97208872; PMID:9055814  
A/Contents: Skin wounds  
A/Accession: JC6197  
A/Molecule type: mRNA  
A/Residues: 1-491 <OKA>  
C/Cross-references: GB:U46034  
C/Comment: This protein is a member of the matrix metalloproteinase family.  
C/Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloproteinase  
C/Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F:52-261/Domain: matrix metalloproteinase homology <MMP>  
F:291-483/Domain: hemopexin repeat homology <HPN>  
F:84-218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status predicted  
F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
F:219/Active site: Glu #status predicted

Query Match 8.3%; Score 180.5; DB 2; Length 491;  
Best Local Similarity 26.7%; Pred. No. 2,8e-07;  
Matches 68; Conservative 22; Mismatches 84; Indels 81; Gaps 13;

Cy 24 TRLPPTSAHGNVASETK---PDPDV-TERGSDGSPDATTLDDNGTMLFFKGEFVWKS 78

Eur. J. Biochem. 235, 187-191, 1996  
 A:Title: Activation of human neutrophil procollagenase by stromelysin 2.  
 A:Reference number: S62608; MUID:96202934; PMID:8631328  
 A:Accession: S62608  
 A:Molecule type: protein  
 A:Residues: 21-39, 'I', '41-47', 'V', '49-122' <KX4>  
 R:Stems: T., Spurlino, J.C.; Smith, D.L.; Rubin, B.  
 submitted to the Brookhaven Protein Data Bank, January 1994  
 A:Reference number: A67078; PDB:1MNC  
 A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 'G', 106-149, 'G', 1  
 R:Stems: T., Spurlino, J.C.; Smith, D.L.; Wahl, R.C.; Ho, T.F.; Goronflah, M.W.; Banks,  
 Nat. Struct. Biol. 1, 119-123, 1994  
 A:Title: Structure of human neutrophil collagenase reveals large S1' specificity pocket.  
 A:Reference number: A58274; MUID:95384762; PMID:7656015  
 A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 'G', 106-149, 'G', 1  
 C:Comment: This protein is more highly glycosylated than interstitial collagenase and is  
 C:Genetics:  
 A:Gene: GDB:IMR8; CLG1  
 A:Cross-references: GDB:128173; OMTM:120355  
 A:Map position: 11q22.2-11q22.3  
 C:Function:  
 A:Description: hydrolyzes collagen types I, II, and III at Gly-Ile sites in collagenous  
 A:Note: cleaves type I collagen most rapidly  
 C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprotei  
 C:Keywords: calcium; extracellular matrix; glycoprotein; hydrolase; metalloproteinase; p  
 F:1-20/Domains: signal sequence #status predicted <SIG>  
 F:21-467/Product: procollagenase #status predicted <PRO>  
 F:21-100/Domains: activation peptide #status experimental <ACT>  
 F:59-262/Domains: matrix metalloproteinase homology <MMP>  
 F:89-96/Region: autoinhibitory  
 F:101-467/Product: neutrophil collagenase #status predicted <MAT>  
 F:203-464/Domains: hemopexin repeat homology <PXN>  
 F:54-73, 112, 119, 204, 246/Binding sites: carboxylate (Asn) (covalent) #status predicted  
 F:73-74/Cleavage site: Asn-Val (autolytic) #status experimental  
 F:84-85/Cleavage site: Asp-Met (autolytic) #status experimental  
 F:91-217, 221, 227/Binding sites: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
 F:99-100/Cleavage site: Phe-Met (autolytic) #status experimental  
 F:167, 169, 182, 195/Binding sites: zinc, noncatalytic (His, Asp, His, His) #status experime  
 F:174, 175, 177, 179, 197, 200/Binding sites: calcium (Asp, Gly, Asn, Ile, Asp, Glu) #status e  
 F:217, 221, 227/Binding sites: zinc, catalytic (His) (active) #status experimental  
 F:218/Active site: Glu #status predicted  
 F:262-263/Cleavage site: Gly-Leu (autolytic) #status experimental  
 F:279-464/Dissulfide bonds: #status predicted  
 Query Match 8.4%; Score 183.5; DB 1; Length 467;  
 Best Local Similarity 24.4%; Pred. No. 1.5e-07;  
 Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;  
 QY 23 ATPLEPTSAHGNVAGETKPPDYTERSCDSGMSFDA-TTLDNGTMLFKGEFVWKS- 79  
 Db 265 SNPIQPT-----GSTPKP-----CDPSLTFTDAITTL--RGEILFFKQRYFWRHPQ 309  
 QY 80 --KMDRELISRWKQNPSPVDAFRQ-GHNSVFLKGDKVVVPEKKKEKPYKLDQ-E 135  
 Db 310 LQREKNMPTSLFWPSLPGIOAAVEDPDRDLIFPKGQVYALSGYDILQCPXDINSG 369  
 QY 136 PFGIPSPUDAVECHRGCAQAEGLVFPFGHGHRNGTGHGNSHTHGPEYWRCSPLVLVSAL 195  
 Db 370 PFPSSVQALDAV-----FTR----- 384  
 QY 196 TSDNGATYAFSGTHYWRDLTSR---DGMHSPIAHQWPGSPASVDAFAFWEEKLYLVQ 252  
 Db 385 -----SKTYFVNQGFMRVYDQRFMERGYPKSLSGAFPGIESKVDVAFQGEHFHFHPSG 439  
 QY 253 TQVYVF 258  
 Db 440 PRVYAF 445

C:Species: Homo sapiens (man)  
 C:Date: 13-Aug-1986 #sequence revision 30-Sep-1992 #text change 08-Dec-2000  
 C:Accession: A37308; S22766; I57620; A00996; D29157; A44518; S06132; B60964; S10595; S53  
 R:Templeton, N.S.; Brown, P.D.; Levy, A.T.; Margulies, I.M.K.; Lotte, L.A.; Settler-Ste  
 Cancer Res. 50, 5431-5437, 1990  
 A:Title: Cloning and characterization of human tumor cell interstitial collagenase.  
 A:Reference number: A37308; MUID:90352587; PMID:2167156  
 A:Accession: A37308  
 A:Molecule type: mRNA  
 A:Residues: 1-469 <TEM>  
 A:Cross-references: GB:X4925; NID:g30125; PIDN:CAA38691.1; PID:g30126  
 R:Brinckerhoff, C.E.; Rudy, P.L.; Auelin, S.D.; Finn, M.E.; White, H.D.  
 J. Clin. Invest. 79, 542-546, 1987  
 A:Title: Molecular cloning of human synovial cell collagenase and selection of a single  
 A:Reference number: S22766; MUID:87109799; PMID:3027129  
 A:Accession: S22766  
 A:Molecule type: DNA  
 A:Residues: 1-63, 65-70 <BRI>  
 A:Cross-references: EMBL:M1596; NID:g180666; PIDN:AAA35700.1; PID:g180667  
 R:Angel, P.; Baumann, I.; Stein, B.; Dellus, H.; Rahmsdorf, H.J.; Herrlich, P.  
 Mol. Cell. Biol. 7, 2256-2266, 1987  
 A:Title: 12-O-tetradecanoyl-phorbol-13-acetate induction of the human collagenase gene i  
 A:Reference number: I57620; MUID:87257941; PMID:3037355  
 A:Accession: I57620  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-35 <RBS>  
 A:Cross-references: GB:M1567; NID:g180668; PIDN:AAA52033.1; PID:g180669  
 R:Goldberg, G.I.; Wilhelm, S.M.; Kronmeyer, A.; Bauer, E.A.; Grant, G.A.; Eiden, A.Z.  
 J. Biol. Chem. 261, 6600-6605, 1986  
 A:Title: Human fibroblast collagenase. Complete primary structure and homology to an onc  
 A:Reference number: A00996; MUID:86196089; PMID:3009463  
 A:Accession: A00996  
 A:Molecule type: mRNA  
 A:Residues: 1-114, 'R', 116-409, 'S', 411-469 <GOL>  
 A:Cross-references: GB:M13509; NID:g180664; PIDN:AAA35699.1; PID:g180665  
 A:Note: part of this sequence was confirmed by protein sequencing of the proenzyme  
 R:Whitham, S.E.; Murphy, G.; Angel, P.; Rahmsdorf, H.J.; Smith, B.J.; Lyons, A.; Harris,  
 Biochem. J. 240, 913-916, 1986  
 A:Title: Comparison of human stromelysin and collagenase by cloning and sequence analysis  
 A:Reference number: A90336; MUID:87156645; PMID:3030290  
 A:Accession: A90336  
 A:Molecule type: mRNA  
 A:Residues: 1-199, 'H', 201-207, 'T', 209-469 <WHI>  
 A:Cross-references: EMBL:X05231; NID:g38266; PIDN:CAA28858.1; PID:g38267  
 A:Note: parts of this sequence, including the amino end of the proenzyme and of the matu  
 R:Birkedal-Hansen, B.; Moore, W.G.I.; Taylor, R.E.; Brown, A.S.; Birkedal-Hansen, H.  
 Biochemistry 27, 6751-6758, 1988  
 A:Title: Monoclonal antibodies to human fibroblast procollagenase. Inhibition of enzymati  
 end of the active enzyme.  
 A:Reference number: A44518; MUID:89062403; PMID:2461732  
 A:Accession: A44518  
 A:Molecule type: protein  
 A:Residues: 270-305 <BIR>  
 R:Clark, I.M.; Dawson, T.E.  
 Biochem. J. 263, 201-206, 1989  
 A:Title: Fragments of human fibroblast collagenase. Purification and characterization.  
 A:Reference number: S06132; MUID:90104231; PMID:2557822  
 A:Accession: S06132  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 100-102, 'P', 104-107, 'XX', 110-112, 270-277, 'X', 279-280, 'X', 282-287 <CLA>  
 R:Lark, M.W.; Walakowite, L.A.; Shah, T.K.; Vammiddleworth, J.; Cameron, P.M.; Lin, T.Y  
 Connect. Tissue Res. 25, 49-65, 1990  
 A:Title: Production and purification of prostromelysin and procollagenase from IL-1 beta-  
 A:Reference number: A60964; MUID:91059606; PMID:2173990  
 A:Accession: A60964  
 A:Molecule type: protein  
 A:Residues: 24-35, 100-108, 270-272, 'X', 274, 'X', 276 <LAB>  
 R:Suzuki, K.; Nagase, H.; Ito, A.; Englund, J.D.; Salvesen, G.  
 Biol. Chem. Hoppe-Seyler 371 (Suppl.), 305-310, 1990  
 A:Title: The role of matrix metalloproteinase 3 in the stepwise activation of human rheun  
 A:Reference number: S10595; MUID:90380300; PMID:2169257

## RESULT 10

KCBBI

interstitial collagenase (EC 3.4.24.7) precursor - rabbit  
 N/Alternate names: fibroblast collagenase; matrix metalloproteinase 1 (MMP1); tissue col  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 18-Jun-1999  
 C/Accession: A27500; B27500; I46694  
 R/Finl, M.E.; Plucinsk, I.M.; Meyer, A.S.; Gross, R.H.; Brinckerhoff, C.E.  
 Biochemistry 26, 6156-6165, 1987  
 A/Title: A gene for rabbit synovial cell collagenase: member of a family of metallopro  
 A/Reference number: A27500; MUID:88077876; PMID:282572  
 A/Accession: A27500  
 A/Molecule type: mRNA  
 A/Residues: 1-468 <FIN>  
 A/Cross-references: GB:M19240  
 A/Accession: B27500  
 A/Molecule type: DNA  
 A/Residues: 1-391;399-468 <F12>  
 A/Cross-references: GB:M17820  
 A/Note: the location of the intron between exons 7 and 8 is approximate  
 R/Finl, M.E.; Austin, S.D.; Holt, P.T.; Rudy, P.L.; Gross, R.H.; White, H.D.; Brinckerhoff  
 Coll. Relat. Res. 6, 239-248, 1986  
 A/Title: Homology between exon-containing portions of rabbit genomic clones for synovial  
 A/Reference number: I46694; MUID:87029174; PMID:3021384  
 A/Accession: I46694  
 A/Status: translated from GB/EMBL/DBSJ  
 A/Molecule type: mRNA  
 A/Residues: 449-468 <F13>  
 A/Cross-references: GB:M25663; NID:G531211; PIDN:AAA31203.1; PID:G531212  
 C/Comment: This enzyme cleaves collagen of types I, II, and III at a Gly-Ile site in th  
 C/Comment: Procollagenase can be activated without removal of the activation peptide. St  
 C/Comment: Procollagenase is found in glycosylated and unglycosylated forms, both of whi  
 C/Genetics:  
 A/Functions: 34/3; 116/2; 166/1; 208/1; 260/1; 299/2; 344/1; 398/1; 433/1  
 A/Description: hydrolyses collagens, in particular types I, II, III, and X, serine, and  
 C/Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprote  
 C/Keywords: calcium; extracellular matrix; fibroblast; glycoprotein; hydrolase; metallo  
 F/1-16/Domain: signal sequence #status predicted <SIG>  
 F/19-468/Product: procollagenase #status predicted <PRO>  
 F/19-48/Domain: activation peptide #status predicted <ACT>  
 F/59-260/Domain: matrix metalloproteinase homology <MMP>  
 F/89-96/Region: autoinhibitory  
 F/99-468/Product: interstitial collagenase #status predicted <NAT>  
 F/271-465/Domain: hemopexin repeat homology <PYN>  
 F/91,217,221,227/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
 F/119,142/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F/217,221,227/Binding site: zinc, catalytic (His) (active) #status predicted  
 F/118/Active site: Glu #status predicted  
 F/277-465/Disulfide bonds: #status predicted

Query Match 8.4%; Score 184.5; DB 1; Length 468;  
 Best Local Similarity 24.1%; Pred. No. 1.2e-07;  
 Matches 54; Conservative 33; Mismatches 80; Indels 57; Gaps 7;

QY 44 PDVTEKSCDGMWFDATLTDNGTMTLFFKGEFVWKSHKMDREL-----ISRRKONFSPVD 98  
 DB 271 PQTPKVCDSKLTDFDAITTI--RGEIMFFKDFRFRANPYSEVELNFISVFWPHLNGIQ 328  
 QY 99 AAFRQGH-NSVFLIKDDKVVVPEPEKKEGPKLLQDEFGIPSPIDANAECHRGCGQAE 157  
 DB 329 AAEEVAHDEILFFKKNKVTVOGQNELPQYPODHSSF----- 367  
 QY 158 GVLFQGHGHRNGTGHGNSTHRGPYWRKSPHLVSLTSDNKGATYAFSGTHYRDL-- 215  
 DB 368 -----GPRPSVNH-----IDAASESDGKTYFFANKNVWYRDEX 402  
 QY 216 -TSRDGHSNPIAHQWPGQPSAVDAAPSWEKKIYLVOGTQVYVF 258  
 DB 403 KESMDAGYPMIEYDPFGIGNKVDAVFKKGDFYFFHGTQYRFX 446

## RESULT 11

KCHUN

neutrophil collagenase (EC 3.4.24.34) precursor [validated] - human  
 N/Alternate names: matrix metalloproteinase 8  
 C/Species: Homo sapiens (man)  
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 23-Mar-2001  
 C/Accession: A37073; A61175; B61175; A36230; S09680; S11026; S19576; S27225; S32527; S626  
 R/Hasty, K.A.; Pourmortabed, T.F.; Goldberg, G.I.; Thompson, J.P.; Spinella, D.G.; Steyer  
 U. Biol. Chem. 265, 11421-11424, 1990  
 A/Title: Human neutrophil collagenase. A distinct gene product with homology to other mat  
 A/Reference number: A37073; MUID:90307647; PMID:2164002  
 A/Accession: A37073  
 A/Molecule type: mRNA  
 A/Residues: 1-467 <HAS>  
 A/Cross-references: GB:J05556; NID:G180617; PIDN:AAA88021.1; PID:G180618  
 R/Devarajan, P.; Mookhtiar, K.; Van Wart, H.; Berliner, N.  
 Blood 77, 2731-2738, 1991  
 A/Title: Structure and expression of the cDNA encoding human neutrophil collagenase.  
 A/Reference number: A61175; MUID:91255696; PMID:1646048  
 A/Accession: A61175  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-31, 'I', 33-86, 'E', 88-467 <DEV>  
 A/Accession: B61175  
 A/Molecule type: protein  
 A/Residues: 263-264, 'X', 266-270, 'X', 272-273, 'X', 275, 'X', 277 <DE2>  
 R/Mallia, S.K.; Mookhtiar, K.A.; Gao, Y.; Brew, K.; Dlozezi, M.; Birkedal-Hansen, H.; V  
 Biochemistry 29, 10628-10634, 1990  
 A/Title: Characterization of 58-kilodalton human neutrophil collagenase: comparison with  
 A/Reference number: A36230; MUID:91104978; PMID:2176876  
 A/Accession: A36230  
 A/Molecule type: protein  
 A/Residues: 'X', 86-87, 'X', 89-90, 'X', 92-97, 'X', 99-111, 'X', 113-120 <MAL>  
 R/Knaeuper, V.; Kraemer, S.; Reinke, H.; Tschesche, H.  
 Eur. J. Biochem. 189, 295-300, 1990  
 A/Title: Characterization and activation of procollagenase from human polymorphonuclear  
 A/Reference number: S09680; MUID:90249372; PMID:2159879  
 A/Accession: S09680  
 A/Molecule type: protein  
 A/Residues: 21-31, 'I', 33-39, 'I', 41-47, 'V', 49-53, 'I', 55-72, 'G', 74-86, 'E', 88-111, 'X', 113-14  
 A/Note: 67-Lys was also found  
 R/Knaeuper, V.; Kraemer, S.; Reinke, H.; Tschesche, H.  
 Biol. Chem. Hoppe-Seyler 371, 733, 1990  
 A/Title: Corrigendum. Partial amino-acid sequence of human PMN leukocyte procollagenase.  
 A/Reference number: S11026; MUID:91000455; PMID:2169766  
 A/Note: original publication was Biol. Chem. Hoppe-Seyler 371 (Suppl.), 295-304, 1990  
 A/Accession: S11026  
 A/Molecule type: protein  
 A/Residues: 21-31, 'I', 33-53, 'I', 55-72, 'G', 74-111, 'X', 113-140;183-203, 'X', 205-209;248-261;  
 A/Note: 87-Glu was also found  
 R/Blaeser, U.; Knaeuper, V.; Osthus, A.; Reinke, H.; Tschesche, H.  
 Eur. J. Biochem. 202, 1223-1230, 1991  
 A/Title: Mercaptal activation of human polymorphonuclear leucocyte procollagenase.  
 A/Reference number: S19576; MUID:92111500; PMID:162606  
 A/Accession: S19576  
 A/Molecule type: protein  
 A/Residues: 69-103 <BL2>  
 R/Blaeser, U.; Tiedel, S.; Reinke, H.; Tschesche, H.  
 FEBS Lett. 313, 59-61, 1992  
 A/Title: Formation of a covalent Hg-Cys-bond during mercaptal activation of PMNL procoll  
 A/Reference number: S27225; MUID:93050220; PMID:1330697  
 A/Accession: S27225  
 A/Molecule type: protein  
 A/Residues: 68-103 <BLA>  
 R/Knaeuper, V.; Osthus, A.; Declercq, Y.A.; Langley, K.E.; Blaeser, U.; Tschesche, H.  
 Biochem. J. 291, 847-854, 1993  
 A/Title: Fragmentation of human polymorphonuclear leucocyte collagenase.  
 A/Reference number: S32527; MUID:93256897; PMID:8489911  
 A/Accession: S32527  
 A/Molecule type: protein  
 A/Residues: 100-112;263-276 <KN3>  
 R/Knaeuper, V.; Murphy, G.; Tschesche, H.

QY 195 LTSDNHGATYAFSGTHYWRDLTSR---DGMHWPJAHQWPGPSAVDAFWSMEKLYLVQ 251  
 Db 399 -----TFYFGDKYKVRPDETRHVMQKGFPRQITDQFPGIEPQVAVLHPEGFYFR 450  
 QY 252 GTQYVVF 258  
 Db 451 GSSQFEF 457

## RESULT 8

KCHS2  
 scromelysin 2 (EC 3.4.24.22) precursor [validated] - human  
 N:Alternate names: matrix metalloproteinase 10 (MMP10); transin-2  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 08-Dec-2000  
 C:Accession: A28816; A47496  
 R:Miller, D.; Quantin, B.; Gesnel, M.C.; Millon-Collard, R.; Abecassis, J.; Breathnach, Biochem. J. 253, 187-192, 1988  
 A:Title: The collagenase gene family in humans consists of at least four members.  
 A:Reference number: A90339; MUID:88339885; PMID:2844164  
 A:Accession: A28816  
 A:Molecule type: mRNA  
 A:Residues: 1-476 <MUI>  
 A:Cross-references: EMBL:X07820; NID:936628; PIDN:CAA30679.1; PID:936629  
 A>Note: mRNA for this protein was detected in several human tumors  
 R:Mindor, L.U.; Grenett, H.; Birkedal-Hansen, B.; Boddien, M.K.; Engler, J.A.; Birkedal-J. Biol. Chem. 268, 17341-17347, 1993  
 A:Title: Cell type-specific regulation of SL-1 and SL-2 genes. Induction of the SL-2 gene  
 A:Reference number: A47496; MUID:93352520; PMID:8349617  
 A:Accession: A47496  
 A:Molecule type: protein  
 A:Residues: 17-33 <MIN>  
 C:Comment: This enzyme degrades various extracellular matrix proteins, including fibronectin  
 C:Genetics:  
 A:Gene: GDB:MMP10; STMY2  
 A:Cross-references: GDB:120392; OMIM:185260  
 A:Map position: 11q22.3-11q23  
 C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprotease  
 C:Keywords: calcium; extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloprotein; signal sequence #status predicted <SIG>  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-476/Product: prostromelysin 2 #status experimental <PRO>  
 F:17-98/Domain: activation peptide #status predicted <ACT>  
 F:59-263/Domain: matrix metalloproteinase homology <MMP>  
 F:89-96/Region: autoinhibitory  
 F:99-476/Product: stromelysin 2 #status predicted <MAT>  
 F:283-476/Domain: hemopexin repeat homology <PXN>  
 F:91,217,221,227/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
 F:119/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:217,221,227/Binding site: zinc, catalytic (His) (active) #status predicted  
 F:218/Active site: Glu #status predicted  
 F:289-476/Disulfide bonds: #status predicted

Query Match 8.7%; Score 189.5; DB 1; Length 476;  
 Best Local Similarity 28.4%; Pred. No. 4.8e-08;  
 Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

QY 25 PLPTSAHGNAVAGETKPDVTERCSQWSPDA-TTLDNDGTMFFKGEVVMKSHKDR 83  
 Db 272 PLVPTK---SVPSGSEMP-----ACVDPALSPDALSTL-RGEYLFKDRYFMRSHNP 321  
 QY 84 E-----LTSERKNPSPVDAAFR-QGNSVFLIKGDKVWV-----YP 120  
 Db 322 EPEPHLISAFWPSLPSYIDAAVEVNSRDTVIFPKNGEMWARGNEVQAGYRGHITLGF 381  
 QY 121 P-----EKK-----EKGYPKLLQDEFPGIPSPIDA 148  
 Db 382 PTIRKIDAAVSGDEKKKTYFFPAADKYWRFDENSGQMEGPFRLADDPGVEPKVDATL- 440  
 QY 149 CHRGECAQAGVLPF 162  
 Db 441 -----QAFGFYF 448

## RESULT 9

## KCHS2

scromelysin 2 (EC 3.4.24.22) precursor - rat

N:Alternate names: matrix metalloproteinase 10 (MMP10); transin-2

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999

C:Accession: B26403; A41775; S26498

R:Breathnach, R.; Macrihan, L.M.; Gesnel, M.C.; Steub, A.; Leroy, P. Nucleic Acids Res. 15, 1139-1151, 1987

A:Title: Sequences coding for part of oncogene-induced transin are highly conserved in a

A:Reference number: A26403; MUID:87146421; PMID:3547333

A:Accession: B26403

A:Molecule type: mRNA

A:Residues: 1-476 <BRE>

A:Cross-references: EMBL:X05083; NID:957388; PIDN:CAA28739.1; PID:957389

A>Note: intron positions were determined by comparison of the cDNA sequence to genomic s

A:Note: mRNA for this protein was expressed in several transformed rat embryo fibroblast

R:Chan, J.C.; Scallion, M.; Zhang, H.Z.; Jia, L.B.; Yu, D.H.; Hung, M.C.; French, M.; Eae

J. Biol. Chem. 267, 1099-1103, 1992

A:Title: Molecular cloning and characterization of v-mos-activated transformation-associ

A:Reference number: A41775; MUID:92112748; PMID:1370458

A:Accession: A41775

A:Molecule type: mRNA

A:Residues: 1-476 <CHA>

A:Cross-references: GB:M65253; NID:9207150; PIDN:AAA42202.1; PID:9207151

A>Note: sequence extracted from NCBI backbone (NCBIP:76184)

R:de Vong, M.W.; Mukherjee, B.B.

Oncogene 7, 109-119, 1992

A:Title: Transformation of normal rat kidney cells by v-K-ras enhances expression of tra

A:Reference number: S26496; MUID:92158347; PMID:1741158

A:Accession: S26498

A:Molecule type: protein

A:Status: preliminary; translation not shown

A:Residues: 31-103, 'L', 241-242, 'TQMEKRP', 251, 'L', 253-254, 'CE', 293-294, 'L', 296 <DEV>

A:Cross-references: EMBL:X64020

C:Genetics:

A:Introns: 35/3; 117/2; 167/1; 209/1; 264/1; 311/2; 356/1; 409/2; 444/1

C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metalloprotease

C:Keywords: calcium; extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloprotein; signal sequence #status predicted <SIG>

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-476/Product: prostromelysin 2 #status predicted <PRO>

F:18-99/Domain: activation peptide #status predicted <ACT>

F:60-264/Domain: matrix metalloproteinase homology <MMP>

F:90-97/Region: autoinhibitory

F:100-476/Product: stromelysin 2 #status predicted <MAT>

F:283-476/Domain: hemopexin repeat homology <PXN>

F:92,218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

F:120/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted

F:219/Active site: Glu #status predicted

F:289-476/Disulfide bonds: #status predicted

Query Match 8.6%; Score 187; DB 1; Length 476;  
 Best Local Similarity 26.4%; Pred. No. 7.8e-08;  
 Matches 52; Conservative 23; Mismatches 48; Indels 74; Gaps 8;

QY 21 AIAFPLPTSAHGNAVAGETKPDVTERCSQWSPDA-TTLDNDGTMFFKGEVVMKSHK 80  
 Db 271 ATTVVPSVS-----PKETPVKCDPALSPFAVTM-LRGEFLFEDRHFMRRTQ 318  
 QY 81 WDRE-----LTSERKNPSPVDAAFRQGH-NSVFLIKGDKVWV----- 118  
 Db 319 WNEPEPHLISAFWPSLPSGIDAAVEANNDKRVILFPGSGQWARGNEVQAGYPKRIHTL 378  
 QY 119 -YYP-----EKK-----EKGYPKLLQDEFPGIPSPIDA 145  
 Db 379 GFPEPTVKIDAAVPEKKEKTYFFVGDKYWRFDTRQLMDGPFRLITDQFPIEPQVDA 438  
 QY 146 AVECHRGCAQAGVLPF 162  
 Db 439 VL-----HAFGFYF 448



A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-445 <RX>  
 A:Cross-references: GB:D50437, NID:g801737, PIDN:BA08928.1, PID:g801738  
 C:Superfamily: hemopexin; hemopexin repeat homology  
 F:42-241/Domain: hemopexin repeat homology <PX1>

Query Match 20.5%; Score 447.5; DB 2; Length 445;  
 Best Local Similarity 28.6%; Pred. No. 9.9e-30;  
 Matches 132; Conservative 68; Mismatches 153; Indels 109; Gaps 16;

```

QY 15 SLCSALATPLPTPTSAHGVAEGETKPPDV-----TERCSGSPFATLTDNGTML 68
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 8 TLCLALSLAAPSNNHKKEDHVQDDEPGQHGLHNGANDRCG-GMEFALAVNEGITY 66
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 69 FRKGEVMSKHKRDLISERKKNP-----SYDAAFRO-----GNNSVLING 113
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 67 FRKGDHLFRGFNDQALSNP---TFPELDEHNLGHVDAFRMSHSDSPAHNDHQFFLD 123
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 114 DKYVVVPRPEKKGKGYRKLQDEPRGI PSPLDAVECHRGCSQAGVLPFG----- 164
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 124 TKVFSTYKHLKEDYFKDISELFPGLPDHDAVBCPTPCANDTILFEDEIYHLMK 183
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 165 -----HGNR-----NGTGHGNSTHHGREY-MRCSF 188
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 184 TKVDEKFKSPNCTGARFYMHHYCFHGHQSKFDPIITGEVQKYPKETRYEMRC-P 242
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 189 H-----LVLSALTDNHCATYAFSGCTHYRLDTRDGMSPRIAHQNP 231
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 243 HFQOKTDEHIEREQSRVHLDAITSDDDGSVAFFGHNH--LSITGDKFHSPTIISAPK 300
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 232 QGSAVDAAPSWEXKLYLVQGTQVYVFLTGKGYTLVSGYKRLKGVKVPFHGILDSVA 291
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 301 ELHSEVDAPFSTEGHLYMKDNEVFYKYGGEPRTHLGGFKPKKEVVG-----IEGPVA 355
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 292 AFICPSSRLHIMAGRLMWLDKSGAQTWTELPMWPH-EKVAGALCMKSLSPNCSAN 350
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 356 AFVCAHDHIAHVYGGTVVDVDKATPRVYVKEGSTALHLKIDIVANC-----GPK----- 405
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 351 GPEGLYIHGPNLYCYSDVEKLNAAKALPQPNVT-SLIGCTH 391
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 406 --GVTAIVIGNHYVQFGSPMIMMAKIMPEQHRVSOGLFGCDH 445
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 6

A49499  
 metalloelastase HME (EC 3.4.24.-) - human  
 C/Species: Homo sapiens (man)  
 C/Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 10-Jul-1998  
 C/Accession: A49499  
 R/Shapiro, S.D.; Kobayashi, D.K.; Ley, T.J.  
 J. Biol. Chem. 268, 23824-23829, 1993  
 A/Title: Cloning and characterization of a unique elastolytic metalloproteinase produced  
 A/Reference number: A49499; MUID:94043200; PMID:8226919  
 A/Accession: A49499  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-470 <SHA>  
 A:Cross-references: GB:L23808, NID:g435969  
 A:Experimental source: alveolar macrophage  
 A/Note: sequence extracted from NCBI backbone (NCBIN:139457, NCIRP:139458)  
 C/Genetic: GDB:MMP12; HME  
 A:Cross-references: GDB:266582; OMIM:601046  
 A/Map position: 11q22.2-11q22.3  
 C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloprote  
 C/Keywords: hydrolase; metalloproteinase; zinc; zymogen  
 F:60-263/Domain: matrix metalloproteinase homology <MMP>  
 F:276-470/Domain: hemopexin repeat homology <PXN>  
 F:93-218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
 F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
 F:219/Active site: Glu #status predicted

Query Match 9.3%; Score 202.5; DB 2; Length 470;  
 Best Local Similarity 27.3%; Pred. No. 3.8e-09;  
 Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

```

QY 30 SAHGVAEGETKPPDVTE--RCSGSPFATLTDNGTMLFPKGEFVW---KSHKWR 83
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 260 SLYGPKEKQENRPNPDNSPCLDPNLSFDAVTVGN-KIFFPKDRFWMVKYSEPKTSV 318
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 84 ELISERKKNPSPVDAAP-OGHNSVPLIKGQVWVYPRPEKKGYPKLQD-EFGGIS 141
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 319 NLISSLMPTLPFGIEAAYEIEARNOVFLFKDKYWLISNLRPEPNVPSIHSFGFPNFK 378
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 142 PLDAVECHRGCSQAGVLPFGQHGRNGTGHGNSTHHGREYRCSPHLVLSALISDNHG 201
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 379 KIDAAV-----FMRPRYR----- 391
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 202 ATYAFSGTHYRLDTRD---GMSWPIAHQPOGSAVDAAPSWEEK-LYLVQGTQY 256
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 392 -TYFVVDNQYMRVDRRQMDPGYRKLITKNFGGICPK-IDAVFYSKNKYVFFQGSNOF 449
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 257 VF 258
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 450 EY 451
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 7

JC6505  
 stromelysin 2 (EC 3.4.24.22) precursor - mouse  
 N/Alternate names: matrix metalloproteinase 10  
 C/Species: Mus musculus (house mouse)  
 C/Date: 16-Oct-1998 #sequence\_revision 16-Oct-1998 #text\_change 21-Jul-2000  
 C/Accession: JC6505  
 R/Madlener, M.; Werner, S.  
 Gene 202, 75-81, 1997  
 A/Title: cDNA cloning and expression of the gene encoding murine stromelysin-2 (MMP-10).  
 A/Reference number: JC6505; MUID:98087420; PMID:9427548  
 A/Accession: JC6505  
 A:Molecule type: mRNA  
 A:Residues: 1-476 <RAD>  
 A:Cross-references: GB:Y1185, NID:g2791311; PIDN:CAA73641.1; PID:g2791312  
 C/Comment: This enzyme degrades various extracellular matrix proteins, including fibronectin  
 C/Genetic: A:Gene: MMP-10  
 C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloprotei  
 C/Keywords: calcium; extracellular matrix; fibroblast; glycoprotein; hydrolase; metallopro  
 F:1-11/Domain: signal sequence #status predicted <SIG>  
 F:18-476/Product: prostromelysin 2 #status predicted <PRO>  
 F:18-99/Domain: activation peptide #status predicted <ACT>  
 F:60-264/Domain: matrix metalloproteinase homology <MMP>  
 F:90-97/Region: autoinhibitory  
 F:100-476/Product: stromelysin 2 #status predicted <MAT>  
 F:283-476/Domain: hemopexin repeat homology <PXN>  
 F:92,218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status p  
 F:210/Binding site: carbohydrate (Aam) (covalent) #status predicted  
 F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
 F:219/Active site: Glu #status predicted  
 F:289-476/Disulfide bonds: #status predicted

Query Match 8.9%; Score 195.5; DB 1; Length 476;  
 Best Local Similarity 26.3%; Pred. No. 1.5e-08;  
 Matches 65; Conservative 26; Mismatches 85; Indels 71; Gaps 10;

```

QY 23 ATPLPTSAHGVAEGETKPPDVTERCSGSPFATLTDNGTMLFRKGEVMSKHKW 81
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 271 ATVPVPLS-----VSPRPETPKCPALSPDSVSTL--RGEVLPFRKDRYFMRSHW 319
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 82 DRE-----LISERKKNPSPVDAAPFROGHN--SVFLIKGQVWVYPRPEKKGYPKLQD- 134
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 320 NPEREHLISATFPTLPDLDAAY--EAHNTDVLFLFKSGQFAVAVRGENEYQAGYRKGIHTL 378
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 135 EPPGISPLDAVECHRGCSQAGVLPFGQHGRNGTGHGNSTHHGREYRCSPHLVLSA 194
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 379 GFPPTVKKIDAAV-----FEKEKKK----- 398
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

A>Title: Identification of an interleukin-6 responsive element and characterization of h  
A/Reference number: A43079; MUID:92287130; PMID:1599480  
A/Accession: A43079  
A/Molecule type: DNA  
A/Residues: 1-14 <NAG>  
A/Cross-references: GB:X60006; NID:g416183; PIDN:CA442621.1; PID:g416184  
R/Nikkliae, H.; Gellin, J.D.; Muller-Eberhard, U.  
Biochemistry 30, 823-829, 1991  
A>Title: Rat hemopexin. Molecular cloning, primary structural characterization, and anal  
A/Reference number: A38399; MUID:91105180; PMID:1988069  
A/Accession: A38399  
A/Molecule type: mRNA  
A/Residues: 15-460 <NIK>  
A/Cross-references: GB:J05306  
R/Muller, D.; Cheng, K.C.; Muller-Eberhard, U.  
Biochem. Biophys. Res. Commun. 155, 622-625, 1988  
A>Title: N-terminal amino acid sequences of the hemopexins from chicken, rat and rabbit.  
A/Reference number: A90148; MUID:88339942; PMID:3421961  
A/Accession: A90148  
A/Molecule type: Protein  
A/Residues: 24-37, 'C', '39-48', 'KW', '51', 'X', '53' <MEL>  
R/Sweris, J.P.; Soula, C.; Sagoc, Y.; Guinaudy, M.J.; Guillemot, J.C.; Ferrara, P.; Dupr  
J. Biol. Chem. 267, 10596-10600, 1992  
A>Title: Hemopexin is synthesized in peripheral nerves but not in central nervous system  
A/Reference number: A38139; MUID:92268104; PMID:1587840  
A/Accession: A38139  
A/Molecule type: Protein  
A/Residues: 24-35, 'X', '37', 'X', '39-43' <SME>  
A/Experimental source: sciatic nerve; skeletal muscle  
A/Note: the amino-terminal sequence of the mature protein was determined  
C/Comment: Hemopexin is a serum glycoprotein that binds heme and transports it to the li  
C/Superfamily: hemopexin; hemopexin repeat homology  
C/Keywords: acute phase; chromoprotein; duplication; glycoprotein; heme; iron; metallopr  
F:1.23/Domain: signal sequence #status predicted <SIG>  
F:2.46/Domain: hemopexin repeat homology predicted <HPX>  
F:4.4-230/Domain: hemopexin repeat homology <PX1>  
F:250-488/Domain: hemopexin repeat homology <PX2>  
F:36,64,186,240,246/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:50-230,148-153,187-199,255-458,364-406,416-433/Diulfide bonds: #status predicted  
F:79,149/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 71.2%; Score 1556.5; DB 1; Length 460;  
Best Local Similarity 63.8%; Pred. No. 9.1e-123;  
Matches 296; Conservative 35; Mismatches 54; Indels 79; Gaps 6;

Qy 1 MARVLGA PVALGLMSLCMSLAIA TPPTSAHGVABEE--TKRDPDVTETCSGMSFDA 58  
Db 1 MARVLGA PVALGLMSLCMSLAIA TPPTSAHGVABEE--TKRDPDVTETCSGMSFDA 58  
Qy 59 TLLDNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 118  
Db 59 TLLDNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 118  
Qy 59 TMDHNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 117  
Db 59 TMDHNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 117  
Qy 119 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 165  
Db 119 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 165  
Qy 118 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 177  
Db 118 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 177  
Qy 166 -----GH-- 167  
Db 166 -----GH-- 167  
Qy 178 ERSWPAVNGCTAALRMLEERYCFQGNKFLRNPVTGEVPPRYPLDADYFISCRGKGK 237  
Db 178 ERSWPAVNGCTAALRMLEERYCFQGNKFLRNPVTGEVPPRYPLDADYFISCRGKGK 237  
Qy 168 -RNGTGHNGSTTHHPPEYWRCSPHLVLSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPI 226  
Db 168 -RNGTGHNGSTTHHPPEYWRCSPHLVLSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPI 226  
Qy 238 LRNGTAGNNGTH--PMHSRCNADRGSLALSNDHGATYA FSGTHYMRLDTSRDGMHSMPI 295  
Db 238 LRNGTAGNNGTH--PMHSRCNADRGSLALSNDHGATYA FSGTHYMRLDTSRDGMHSMPI 295  
Qy 227 AHQWPGPSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 286  
Db 227 AHQWPGPSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 286  
Qy 296 AHHPQGPSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 355  
Db 296 AHHPQGPSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 355  
Qy 287 DSVDAAATCGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 346  
Db 287 DSVDAAATCGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 346  
Qy 356 DTIDAAFSCKSSSKLYVTSGRRLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 415  
Db 356 DTIDAAFSCKSSSKLYVTSGRRLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 415

Qy 347 CSANGPGLYLHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 390  
Db 416 CSANGPGLYLHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 390  
Qy 416 CSANGPGLYLHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 390  
Db 416 CSANGPGLYLHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 390

RESULT 4  
A55486  
hyaluronidase - pig  
N/Alternate names: hemopexin  
C/Species: Sus scrofa domestica (domestic pig)  
C/Date: 03-Mar-1995 #sequence\_revision 03-Mar-1995 #text\_change 20-Aug-1999  
C/Accession: A55486  
R/Zhu, L.; Hope, T.J.; Hall, J.; Davies, A.; Stern, M.; Muller-Eberhard, U.; Stern, R.;  
J. Biol. Chem. 269, 32092-32097, 1994  
A>Title: Molecular cloning of a mammalian hyaluronidase reveals identity with hemopexin.  
A/Reference number: A55486; MUID:95096047; PMID:7798203  
A/Accession: A55486  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-459 <ZHU>  
A/Cross-references: GB:U14751; NID:g541627; PIDN:AAC48457.1; PID:g541628  
C/Superfamily: hemopexin; hemopexin repeat homology  
F:46-232/Domain: hemopexin repeat homology <PX1>  
F:245-453/Domain: hemopexin repeat homology <PX2>

Query Match 70.7%; Score 1544; DB 2; Length 459;  
Best Local Similarity 63.4%; Pred. No. 1e-121;  
Matches 289; Conservative 33; Mismatches 64; Indels 70; Gaps 5;

Qy 1 MARVLGA PVALGLMSLCMSLAIA TPPTSAHGVABEE--TKRDPDVTETCSGMSFDA 58  
Db 1 MARVLGA PVALGLMSLCMSLAIA TPPTSAHGVABEE--TKRDPDVTETCSGMSFDA 58  
Qy 59 TLLDNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 118  
Db 59 TLLDNGTMLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 118  
Qy 61 STLDEHGMPLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 120  
Db 61 STLDEHGMPLFFKGEFVWKSHKMDRELISERKNFSPVDAAFRQGHNSVFLIKGDKVM 120  
Qy 119 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 164  
Db 119 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 164  
Qy 121 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 179  
Db 121 YPEKKEKGYPKLQDEFPGLPSPFLDAVACHRECOAGVLPFGH----- 179  
Qy 165 -----GH-- 173  
Db 165 -----GH-- 173  
Qy 180 ERLWPAVNGCTAALRMLEERYCFQGNKFLRNPVTGEVPPRYPLDADYFISCRGKGK 239  
Db 180 ERLWPAVNGCTAALRMLEERYCFQGNKFLRNPVTGEVPPRYPLDADYFISCRGKGK 239  
Qy 174 GNSGTHHPPEYWRCSPHLVLSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPIAHQWPG 233  
Db 174 GNSGTHHPPEYWRCSPHLVLSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPIAHQWPG 233  
Qy 240 RNATHRGDD--RCSPLDVLTSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPIAHQWPG 297  
Db 240 RNATHRGDD--RCSPLDVLTSALTSNDHGATYA FSGTHYMRLDTSRDGMHSMPIAHQWPG 297  
Qy 234 PSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 293  
Db 234 PSAVDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 293  
Qy 298 PSANDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 357  
Db 298 PSANDAAFSBEKLYLVQGTQVYVFLTKGGYTVLVSGYPKRLEKEVGTPHGIL 357  
Qy 294 ICPGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 353  
Db 294 ICPGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 353  
Qy 358 VCPGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 417  
Db 358 VCPGSSSLHIMAGGRLLMWLDLKSQAQNTWTELPMPEHKKVDAALCMESLGPNS 417  
Qy 354 LYLIHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 389  
Db 354 LYLIHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 389  
Qy 418 LYTIHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 453  
Db 418 LYTIHGNLYCYSDVEKLNAAKALPOPONTSLIGCT 453

RESULT 5  
I50485  
Warm temperature acclimation-related 65-kDa protein, Wap65 - goldfish  
C/Species: Carassius auratus (goldfish)  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C/Accession: I50485  
R/Kikuchi, K.; Yamashita, M.; Watabe, S.; Aida, K.  
J. Biol. Chem. 270, 17087-17092, 1995  
A>Title: The warm temperature acclimation-related 65-kDa protein, Wap65, in goldfish  
A/Reference number: I50485; MUID:95340486; PMID:7615502  
A/Accession: I50485

A:Residues: 24-49;62-67;178-192;236-250;443-457 <TA2>  
 C:Comment: Hemopexin is a serum glycoprotein that binds heme and transports it to the 11  
 C:Genetics:  
 A:Gene: GDB:HPX  
 A:Cross-references: GDB:1120054; OMIM:142290  
 A:Map position: 11p15.4-11p15.4  
 A:Introns: 28/2; 48/1; 72/1; 112/3; 164/1; 235/1; 279/1; 322/3; 377/1  
 C:Superfamily: hemopexin; hemopexin repeat homology  
 C:Keywords: acute phase; duplication; glycoprotein; heme binding; plasma  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-462/Product: hemopexin #status experimental <HPX>  
 F:44-231/Domain: hemopexin repeat homology <PX2>  
 F:252-460/Domain: hemopexin repeat homology <PX2>  
 F:50-231/Binding site: carbohydrate (Thr) (covalent) #status experimental  
 F:50-231,149-154,188-200,257-460,418-435/Disulfide bonds: #status experimental  
 F:44,187,240,246,453/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 97.9%; Score 2139.5; DB 1; Length 462;  
 Best Local Similarity 84.6%; Pred. No. 1,1e-171;  
 Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;

QY	1	MARVLGAPVALGWSLCSLAIATPLPPTSAHGNVAEGETKPPDPTTERCSDDGMSFDATT	60
DB	1	MARVLGAPVALGWSLCSLAIATPLPPTSAHGNVAEGETKPPDPTTERCSDDGMSFDATT	60
QY	61	LDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	120
DB	61	LDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	120
QY	121	PEKKEKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPO-----GHGHRN	163
DB	121	PEKKEKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPO-----GHGHRN	163
QY	164	-----GHGHRN	169
DB	164	-----GHGHRN	169
QY	170	GTGCHGNSHTHGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	229
DB	170	GTGCHGNSHTHGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	229
QY	230	WPGGSAVDAAFSWBEKLYLVGCTQYVFLTKGGYTLVSGYPRLEKEVTPHGIILDSV	289
DB	230	WPGGSAVDAAFSWBEKLYLVGCTQYVFLTKGGYTLVSGYPRLEKEVTPHGIILDSV	289
QY	290	DAAFICPGSSRLHIMGRRLMWLDLKSQAQTWTELPWHEKVDGALCMEKSLGPNSCSA	349
DB	290	DAAFICPGSSRLHIMGRRLMWLDLKSQAQTWTELPWHEKVDGALCMEKSLGPNSCSA	349
QY	350	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	391
DB	350	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	391
QY	421	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	462
DB	421	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	462

RESULT 2  
 OORR  
 N:Alternate names: beta-1b-glycoprotein  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 21-Sep-1993 #sequence, revision 13-Jan-1995 #text\_change 03-Mar-2000  
 C:Accession: A46006; B46006; A61426; B31514  
 R:Morgan, W.T.; Musteer, P.; Tatum, F.; Kao, S.M.; Alam, J.; Smith, A.  
 J. Biol. Chem. 268, 6256-6262, 1993  
 A:Title: Identification of the histidine residues of hemopexin that coordinate with heme  
 A:Reference number: A46006; MOID:93203213; PMID:7681064  
 A:Accession: A46006  
 A:Molecule type: mRNA  
 A:Residues: 1-459 <MOR>  
 A:Note: sequence extracted from NCBI backbone (NCBIN:127918, NCBI:P.127919)  
 A:Accession: B46006  
 A:Molecule type: protein  
 A:Residues: 26-33;104-108;239-242, 'X', 243-244 <MO2>  
 A:Note: residues 42-46 are shown as 'TKPEA' in the Fig. 2 alignment, with no explanation

R:Muster, P.; Tatum, F.; Smith, A.; Morgan, W.T.  
 J. Protein Chem. 10, 123-128, 1991  
 A:Title: Further characterization of structural determinants of rabbit hemopexin function  
 A:Reference number: A61426; MOID:91273754; PMID:2054057  
 A:Accession: A61426  
 A:Molecule type: protein  
 A:Residues: 26-41, 'TKPEA', 47-51 <MUS>  
 R:Wellner, D.; Cheng, K.C.; Muller-Eberhard, U.  
 Biochem. Biophys. Res. Commun. 155, 622-625, 1988  
 A:Title: N-terminal amino acid sequences of the hemopexins from chicken, rat and rabbit.  
 A:Reference number: A90146; MOID:88339942; PMID:3421961  
 A:Accession: B31514  
 A:Molecule type: protein  
 A:Residues: 26-33, 'X', 35-41, 'TKPEA', 47-51, 'W', 53 <MB>  
 C:Comment: Hemopexin is a serum glycoprotein that binds heme and transports it to the 11  
 C:Keywords: acute phase; hemopexin; hemopexin repeat homology  
 C:Superfamily: hemopexin; hemopexin repeat homology  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-459/Product: hemopexin #status experimental <HPX>  
 F:46-232/Domain: hemopexin repeat homology <PX2>  
 F:249-457/Domain: hemopexin repeat homology <PX2>  
 F:24/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:52-232,150-155,189-201,254-457,363-405,415-432/Disulfide bonds: #status predicted  
 F:56,188,243/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:51,151/Binding site: heme iron (His) (axial ligands) #status experimental

Query Match 77.2%; Score 1686; DB 1; Length 459;  
 Best Local Similarity 67.3%; Pred. No. 1,2e-133;  
 Matches 311; Conservative 28; Mismatches 49; Indels 74; Gaps 4;

QY	1	MARVLGAPVALGWSLCSLAIATPLPPTSAHGNVAEGETKPPDPTTERCSDDGMSFDPA	58
DB	1	MARVLGAPVALGWSLCSLAIATPLPPTSAHGNVAEGETKPPDPTTERCSDDGMSFDPA	60
QY	59	TTLDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	118
DB	59	TTLDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	118
QY	61	TTLDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	119
DB	61	TTLDDNGTMLFFKGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	119
QY	119	YPEKKEKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPOH-----	165
DB	119	YPEKKEKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPOH-----	165
QY	120	YPPDKKDKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPOH-----GHGHRN	179
DB	120	YPPDKKDKGYPKLLDDEFPGIPSPDLAAVECHRGECQAAGVLFPOH-----GHGHRN	179
QY	166	-----GHGHRN	169
DB	166	-----GHGHRN	169
QY	170	GTGCHGNSHTHGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	229
DB	170	GTGCHGNSHTHGEFVWVSKHMDRELISERKWNFPSPVDAAFROGHNSVFLIKGDKVWVP	229
QY	230	WPGGSAVDAAFSWBEKLYLVGCTQYVFLTKGGYTLVSGYPRLEKEVTPHGIILDSV	289
DB	230	WPGGSAVDAAFSWBEKLYLVGCTQYVFLTKGGYTLVSGYPRLEKEVTPHGIILDSV	289
QY	290	DAAFICPGSSRLHIMGRRLMWLDLKSQAQTWTELPWHEKVDGALCMEKSLGPNSCSA	349
DB	290	DAAFICPGSSRLHIMGRRLMWLDLKSQAQTWTELPWHEKVDGALCMEKSLGPNSCSA	349
QY	350	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	391
DB	350	NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	391
QY	418	SGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	459
DB	418	SGPGLYLHIGPNLYCYSDVEKLNAAKALPOPOVNTSLIGCTH	459

RESULT 3  
 OORR  
 N:Alternate names: beta-1b-glycoprotein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 28-Jun-1991 #sequence, revision 13-Jan-1995 #text\_change 03-Mar-2000  
 C:Accession: A43073; A58399; A31514; A38139  
 R:Nagae, Y.; Muller-Eberhard, U.  
 Biochem. Biophys. Res. Commun. 185, 420-429, 1992

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 16, 2003, 05:47:17 ; Search time 34 Seconds  
(without alignments)  
1105.940 Million cell updates/sec

Title: US-09-900-448-2

Perfect score: 2185  
Sequence: 1 MARVLGAPVALGLWSLMSL.....NAAKALPQPQNTVSLGCTH 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2139.5	97.9	462	1	OOBU hemopepin precursor
2	1686	77.2	459	1	OOBU hemopepin precursor
3	1556.5	71.2	460	1	COBT hyaluronidase - pi
4	1544	70.7	459	2	Wam temperature a
5	447.5	20.5	445	2	A50485 metalloelastase HM
6	202.5	9.3	470	2	A49499 stromelysin 2 (EC
7	195.5	8.9	476	1	JC6505 stromelysin 2 (EC
8	189.5	8.7	476	1	KCHUS2 stromelysin 2 (EC
9	187	8.6	476	1	KCRB2 stromelysin 2 (EC
10	184.5	8.4	468	1	KCHUN1 interstitial colla
11	183.5	8.4	467	1	KCHUN1 interstitial colla
12	181.5	8.3	469	1	KCHUN1 interstitial colla
13	181.5	8.3	492	2	A44399 stromelysin 3 (EC
14	180.5	8.3	491	2	JC6197 stromelysin 3 (EC
15	179.5	8.2	488	2	S13423 stromelysin 3 (EC
16	177	8.1	475	1	KCRTH1 stromelysin 1 (EC
17	177	8.1	477	1	KCHUS1 stromelysin 1 (EC
18	176	8.1	471	1	A53711 collagenase 3 (EC
19	173	7.9	477	1	KCMS51 stromelysin 1 (EC
20	172	7.9	669	2	I38029 matrix metalloprot
21	170.5	7.8	478	1	SGUUV1 vitronectin precur
22	169.5	7.8	477	1	I51645 stromelysin 3 (EC
23	169.5	7.8	521	2	T37252 probable matrix me
24	167.5	7.7	508	2	JC5082 matrix metalloprot
25	166.5	7.6	466	2	A23685 interstitial colla
26	166	7.6	469	1	KCRPT1 interstitial colla
27	166	7.6	476	1	SGMSV vitronectin precur
28	166	7.6	478	1	KCRB51 stromelysin 1 (EC
29	158	7.2	459	2	JC5139 vitronectin precur

30	154.5	7.1	472	2	S29243 interstitial colla
31	151	6.9	469	1	KCB01 interstitial colla
32	150.5	6.9	582	2	I38028 matrix metalloprot
33	149.5	6.8	475	2	A38340 66k glycoprotein p
34	149.5	6.8	582	2	I84471 matrix metalloprot
35	147.5	6.8	462	2	A42401 macropinase elasta
36	143.5	6.6	663	1	S46492 gelatinase A (EC 3
37	142.5	6.5	662	2	S70365 gelatinase A (EC 3
38	142	6.5	662	2	A42496 gelatinase A (EC 3
39	142	6.5	662	2	S34780 gelatinase A (EC 3
40	140.5	6.4	660	1	A28153 gelatinase A (EC 3
41	139	6.4	484	2	A40774 phosphocholine-bin
42	135	6.2	707	1	A34458 gelatinase B (EC 3
43	133.5	6.1	582	2	I48673 matrix metalloprot
44	124.5	5.7	616	2	JC7776 matrix metalloprot
45	124	5.7	483	2	JC5743 matrix metalloprot

ALIGNMENTS

RESULT 1  
hemopepin precursor [validated] - human  
N:Alternate names: beta-1B-glycoprotein  
C:Species: Homo sapiens (man)  
C>Date: 28-Aug-1985 #sequence revision 02-Jul-1996 #text change 08-Dec-2000  
C:Accession: I56456; I54212; A93566; A94069; A93328; A43791; A03263  
R:Altitude, F.; Poli, V.; Restagno, G.; Sillengo, L.  
J. Mol. Evol. 27, 102-108, 1988  
A>Title: Structure of the human hemopepin gene and evidence for intron-mediated evolution  
A:Reference number: I56456; MUID:88316972; PMID:2842511  
A:Accession: I56456  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-462 <RES>  
A:Cross-references: GB:M36803; NID:G184495; PIDN:AAA5878.1; PID:G184497  
R:Altitude, F.; Poli, V.; Restagno, G.; Argos, P.; Cortese, R.; Sillengo, L.  
Nucleic Acids Res. 13, 3841-3859, 1985  
A>Title: The primary structure of human hemopepin deduced from cDNA sequence: evidence f  
A:Reference number: A93566; MUID:85242073; PMID:2969777  
A:Accession: A93566  
A:Molecule type: mRNA  
A:Residues: 2-462 <RES>  
A:Cross-references: GB:J03048; NID:G184487; PIDN:AAA52704.1; PID:G386789  
R:Altitude, F.; Poli, V.; Restagno, G.; Argos, P.; Cortese, R.; Sillengo, L.  
Nucleic Acids Res. 13, 3841-3859, 1985  
A>Title: Complete amino acid sequence of human hemopepin, the heme-binding protein of se  
A:Reference number: A94069; MUID:85113173; PMID:3855550  
A:Accession: A94069  
A:Molecule type: protein  
A:Residues: 24-462 <TAK>  
R:Frantlikova, V.; Borvak, J.; Kluh, I.; Moravsek, L.  
FEBS Lett. 178, 213-216, 1984  
A>Title: Amino acid sequence of the N-terminal region of human hemopepin.  
A:Reference number: A91328; MUID:85076955; PMID:6510521  
A:Accession: A91328  
A:Molecule type: protein  
A:Residues: 24-255 <PRA>  
R:Takahashi, N.; Takahashi, Y.; Putnam, F.W.  
Proc. Natl. Acad. Sci. U.S.A. 81, 2021-2025, 1984  
A>Title: Structure of human hemopepin: O-glycosyl and N-glycosyl sites and unusual clust  
A:Reference number: A43791; MUID:84193947; PMID:6371807  
A:Accession: A43791  
A:Molecule type: protein

DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.  
KM Hydrolase; Metalloprotease; zinc; Zymogen; Calcium;  
KM Collagen degradation; Extracellular matrix; Signal; 3D-structure.  
FT SIGNAL 1 35  
FT PROPEP 36 101  
FT CHAIN 102 492  
FT DOMAIN 295 487  
FT SITE 84 84  
FT METAL 219 219  
FT ACT SITE 220 220  
FT METAL 223 223  
FT METAL 229 229  
FT DISULFID 298 484  
FT STRAND 102 105  
FT TURN 106 107  
FT STRAND 108 111  
FT STRAND 113 118  
FT HELIX 127 142  
FT TURN 143 144  
FT STRAND 148 151  
FT STRAND 159 164  
FT STRAND 182 185  
FT TURN 188 189  
FT STRAND 193 198  
FT TURN 199 200  
FT STRAND 203 204  
FT TURN 208 209  
FT STRAND 211 212  
FT HELIX 213 224  
FT TURN 225 226  
FT TURN 233 234  
FT TURN 236 237  
FT HELIX 251 261  
SQ SEQUENCE 492 AA; 55441 MW; B54E260E4AB3D7C3 CRC64;  
  
Query Match 8.4; Score 184.5; DB 1; Length 492;  
Best Local Similarity 25.2; Pred. No. 2.9e-08;  
Matches 70; Conservative 27; Mismatches 88; Indels 93; Gaps 13;  
  
QY 20 LAIATPIPTSAH-----GNVAGETKPDVDYTERCSDGMSFDATTLDNGTMLFFRGER 74  
DB 266 MAPTSPAPVLTSSQAGTDTNEIALLEPTPPVCET-----SFDVVS-TIRGELFFRQGF 319  
QY 75 VMKSHKWDRE-----LISERKNFPPSPVDAAFRQGHNSVFLIKGDKVWVYPPKEKEGYR 129  
DB 320 VMRLRSGRLOPGYPALASRMQGLPSVDAAFEBAQGIWFFQAGQYVYDGEKPVLGPA 379  
QY 130 KLIQDEFPGLPSPLDAAVECHRGECQAEVLFQGHGRNGTGHGNSTHHGPEYMRCSPH 189  
DB 380 PLSKLGIQG--SPVHALV-----WGPR----- 400  
QY 190 LVLSALTSNDHGATYASGTHYWRLD--TSR-DGMHSWPIAHQ--WPGPSAVDAAFSW 243  
DB 401 -----KNKIYFRGGDYWFHPRTORVDN---PVRSTDRGVPSSEIDAF-- 444  
QY 244 EEKLYLVQGTQVYVFLTKG-----GYLVSQYPR 272  
DB 445 -----QDAGYAYFLRGHLVWKFDPVKVLEGFPR 475

Search completed: December 16, 2003, 06:48:52  
Job time : 30 secs

CC -1- FUNCTION: CLEAVES COLLAGENS OF TYPES I, II, AND III AT ONE SITE IN  
 CC THE HELICAL DOMAIN. ALSO CLEAVES COLLAGENS OF TYPES VII AND X.  
 CC -1- CATALYTIC ACTIVITY: Cleaves preferentially one bond in native  
 CC collagen. Cleavage of the triple helix of collagen at about three-  
 CC quarters of the length of the molecule from the N-terminus, at  
 CC 775-Gly-Ile-776 in the alpha-1(I) chain. Cleaves synthetic  
 CC substrates and alpha-macroglobulins at bonds where Pl<sup>1</sup> is a  
 CC hydrophobic residue.  
 CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.  
 CC -1- ENZYME REGULATION: CAN BE ACTIVATED WITHOUT REMOVAL OF THE  
 CC ACTIVATION PEPTIDE.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.  
 CC -1- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M17823; AAB88016.1; -  
 DR EMBL; M17820; AAB88016.1; JOINED.  
 DR EMBL; M17821; AAB88016.1; JOINED.  
 DR EMBL; M17822; AAB88016.1; JOINED.  
 DR EMBL; M19240; AAB88016.1; JOINED.  
 DR EMBL; M25663; AAA31203.1; -  
 DR PIR; A27500; KCRBI.  
 DR HSSP; P03956; ICGI.  
 DR MEROPS; M10.001; -  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR001818; Matrxin.  
 DR InterPro; IPR006026; Nzn\_MTPeptidase.  
 DR InterPro; IPR006025; Zn\_MTPeptidase.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR Pfam; PF03933; Peptidase\_M10\_N; 1.  
 DR PRINTS; PR00138; MATR1XIN.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00235; ZNMC; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.  
 KM Hydroxylase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;  
 KW Collagen degradation; Extracellular matrix; Signal.  
 FT STGNAL 1 18  
 FT PROPEP 19 98  
 FT CHAIN 99 468  
 FT DOMAIN 274 468  
 FT SITE 91 91  
 FT METAL 217 217  
 FT ACT\_SITE 218 218  
 FT METAL 221 221  
 FT METAL 227 227  
 FT CARBOHYD 119 119  
 FT DISULFID 277 465  
 SQ SEQUENCE 468 AA; 53739 MW; DA90538919952B8C CRC64;  
 Query Match 8.4%; Score 184.5; DB 1; Length 468;  
 Best Local Similarity 24.1%; Pred. No. 2.7e-08;  
 Matches 54; Conservative 33; Mismatches 80; Indels 57; Gaps 7;

DB 368 -----GPRSVNH-----IDAASEDGTGTYFPVANKYRYDEY 402  
 QY 216 -TSRDGWSHPAIAHQWPGSAVDAAPSAWSEKLYVGOTGVYV 258  
 DB 403 KRSDIAGYPMIEYDFPGIGKNDVAVPKDGFYFFPHGTROYKF 446  
 RESULT 15  
 ID MM11\_MOUSE STANDARD; PRT; 492 AA.  
 AC Q02853;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Stromelysin-3 precursor (EC 3.4.24.-) (Matrix metalloproteinase-11)  
 DE (MMP-11) (ST) (SL-3).  
 GN MMP11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=93054930; Pubmed=1429845;  
 RA Lefebvre O., Wolf C., Limacher J.M., Rutin P., Wendling C.,  
 RA Lemer M., Basset P., Rio M.C.;  
 RT "The breast cancer-associated stromelysin-3 gene is expressed during  
 RT mouse mammary gland apoptosis";  
 RL J. Cell Biol. 119:997-1002(1992).  
 RN [2]  
 RP REVISIONS.  
 RA Lefebvre O.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 102-265.  
 RX MEDLINE=21154142; Pubmed=11254383;  
 RA Gall A.-L., Ruff M., Kannan R., Cunnase P., Viotakis A., Dive V.,  
 RA Rio M.-C., Basset P., Moras D.;  
 RT "Crystal structure of the stromelysin-3 (MMP-11) catalytic domain  
 RT complexed with a phosphinic inhibitor mimicking the  
 RT transition-state";  
 RL J. Mol. Biol. 307:577-586(2001).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE PROGRESSION OF  
 CC EPITHELIAL MALIGNANCIES.  
 CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE MAMMARY GLAND  
 CC DURING APOPTOSIS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.  
 CC -1- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z12604; CA78248.1; -  
 DR PDB; 1HVS; 28-MAR-01.  
 DR MEROPS; M10.007; -  
 DR MGD; MGI:97008; MMP11.  
 DR InterPro; IPR000585; Hemopexin.  
 DR InterPro; IPR001818; Matrxin.  
 DR InterPro; IPR006026; Nzn\_MTPeptidase.  
 DR InterPro; IPR006025; Zn\_MTPeptidase.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR PRINTS; PR00138; MATR1XIN.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00235; ZNMC; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

```

Db 272 PLVPTK---SVPSGSEMP-----AKCDPALSPDAISLT--RGSEYLFKDRYFRSHMNP 321
Qy 84 E-----LISERKWKPFSPVDAAFR-QGHSNFKLKGDKWV-----YF 120
Db 322 EPEFHLSAIFWPLPSYLDAAVEVNSRDYFIRKGFEMAIRNGENVOAGYPRGIIHTLGF 361
Qy 121 P-----EKK-----EKGYPKLLDPEFGIPSPDLAAVE 148
Db 382 PIRIKDAVSDSEKKKTYFPADKWRFPDENSQSGMEGFPRLADDFPVEKRVDAVL- 440
Qy 149 CHRGECAEGVLFP 162
Db 441 -----QAGFFYF 448

RESULT 13
MM10_RAT
ID MM10_RAT STANDARD; PRT; 476 AA.
AC P07152;
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Stromelysin-2 precursor (EC 3.4.24.22) (Matrix metalloproteinase-10) (MMP-10) (Transin-2) (SL-2) (Transformation-associated protein 34A).
GN MMP10.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8716421; PubMed=3547333;
RA Breachmach R., Matriasian L.M., Gesnel M.-C., Staud A., Leroy P.;
RT "Sequences coding for part of oncogene-induced transin are highly
RT conserved in a related rat gene.";
RL Nucleic Acids Res. 15:1139-1151(1987).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112748; PubMed=1370458;
RA Chan J.C., Scanlon M., Zhang H.Z., Jia L.B., Yu D., Hung M.C.,
RA French M., Eastman E.M.;
RT "Molecular cloning and characterization of v-mos-activated
RT transformation-associated proteins.";
RL J. Biol. Chem. 267:1099-1103(1992).
CC - FUNCTION: CAN DEGRADE FIBRONECTIN, GELATIN OF TYPE I, III, IV,
CC AND V; WEAKLY COLLAGENS III, IV, AND V. ACTIVATES PROCOLLAGENASE.
CC - CATALYTIC ACTIVITY: Similar to stromelysin 1, but action on
CC collagen types III, IV and V is weak.
CC - COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC - SIMILARITY: Contains 1 hemopexin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05083; CA28739.1; -
DR EMBL: M65253; AAA42202.1; -
DR FTR; B26403; KCRTS2.
DR HSSP; P08254; ISLM.
DR MEROPS; M10.006; -
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001818; Matrixin.
DR InterPro: IPR006026; Zn_MTPeptide.
DR InterPro: IPR006025; Zn_MTPeptide.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase_M10; 1.
DR Pfam; PF03933; Peptidase_M10_N; 1.

```

```

DR PRINTS; PRO0138; MATRIXIN.
DR SMART; SM00120; HK; 3.
DR PROSITE; PS00235; ZINC; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal.
FT SIGNAL 1 17 PROBABLE.
FT PROPEP 18 99 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 100 476 STROMELYSIN-2.
FT DOMAIN 286 476 HEMOPEXIN-LIKE.
FT SITE 92 92 CYSTEINE SWITCH (BY SIMILARITY).
FT METAL 218 218 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 219 219 BY SIMILARITY.
FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 289 476 BY SIMILARITY.
SQ SEQUENCE 476 AA; 54221 MW; B556BFB1D8BAYEE CRC64;

Query Match 8.6%; Score 187; DB 1; Length 476;
Best Local Similarity 26.4%; Pred. No. 1.7e-08;
Matches 52; Conservative 23; Mismatches 48; Indels 74; Gaps 8;

Qy 21 AIAITPLPTSAHGNVAEGETKPPDVTETRCSDGMSFDATTLDDNGTMLEFKGEFVWKS HK 80
Db 271 ATVPVPSVS-----PKETPVKCDPALSPDAVTM-LRGSEYLFKDRYFRSHMNP 318
Qy 81 WDRE---LISERKWKPFSPVDAAFRQGH-NSVFLIKGDKWV----- 118
Db 319 WNEPEFHLSAIFWPLPSYLDAAVEANNNKRVLIKFGSQFNAVNRGENVOAGYPRKIIHTL 378
Qy 119 -YRP-----EKK-----EKGYPKLLDPEFGIPSPDLA 145
Db 379 GFPPYTKKIDAAVEEKKKTYFPVGDKWRPDETQLMDKGFRLITDDFPGEIQQVDA 438
Qy 146 AVECHRGECAGVLFP 162
Db 439 VL-----HAGFFYF 448

RESULT 14
MM01_RABIT
ID MM01_RABIT STANDARD; PRT; 468 AA.
AC P13453;
DT 01-JAN-1990 (Rel. 13, Created).
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interstitial collagenase precursor (EC 3.4.24.7) (Matrix
DE metalloproteinase-1) (MMP-1).
GN MMP1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Synovial cell;
RX MEDLINE=88077876; PubMed=2825772;
RA Fani M.E., Plucinska I.M., Mayer A.S., Gross R.H., Brinckerhoff C.E.;
RT "A gene for rabbit synovial cell collagenase: member of a family of
RT metalloproteinases that degrade the connective tissue matrix.";
RL Biochemistry 26:6156-6165(1987).
RN [2]
RP SEQUENCE OF 449-468 FROM N.A.
RX STRAIN=New Zealand white;
RX MEDLINE=87029174; PubMed=3021384;
RA Fani M.E., Austin S.D., Holt P.T., Ruby P.L., Gross R.H., White H.D.,
RA Brinckerhoff C.E.;
RT "Homology between exon-containing portions of rabbit genomic clones
RT for synovial cell collagenase and human foreskin and synovial cell
RT mRNA.";
RL Coll. Relat. Res. 6:239-248(1986).

```



```

FT TURN 123 124
FT HELIX 127 144
FT STRAND 148 151
FT STRAND 159 164
FT STRAND 182 184
FT TURN 190 193
FT STRAND 195 198
FT TURN 199 200
FT STRAND 203 204
FT STRAND 210 211
FT HELIX 212 224
FT TURN 225 225
FT TURN 232 233
FT TURN 235 236
FT HELIX 253 263
SQ SEQUENCE 477 AA; 54190 MW; 361CE1427E09A272 CRC64;

Query Match 8.9%; Score 195.5; DB 1; Length 477;
Beet Local Similarity 27.3%; Pred. No. 3.2e-09;
Matches 67; Conservative 24; Mismatches 89; Indels 65; Gaps 11;

OY 27 PPTSAHGVABGETPD-PDVTGRCSDDGWSFDA-TTLDNGTMLPFKGEFVKSHKMDRE 84
DB 266 PPASDSEVSEPPSPAPRTIAMCDPALSPDAISTL--RGEILFFKORYFWR--KTRRT 321
OY 85 L-----ISERWKNPSPVDAAFR-QGHNSVFLKGDVWVYPPKKEKGYPKLQD-EF 136
DB 322 LVPEHFPISSFWPSPSPGIDAAVEYTSRDSVIFPKGNFMAIRGNEDQAGYRGHITIGF 381
OY 137 PGISPLDAVACHRGECQAGVLFPGQGHNRNGTGHGNSHTHGPEYKRCSPHLVLSALT 196
DB 382 PPTVAKKIDAI-----FDEKKEK----- 399
OY 197 SDNHGATYAFSGTHYRLDTSRDGWH-SMP--IAHQWQGSAPVADAFSWEEKLYVQGT 253
DB 400 -----TYFVEDKTKWRDEKQSGMEPGYKQIADDPGIDSKLDAFESGFFYFSGS 453
OY 254 QVYVF 258
DB 454 SQFEF 458

RESULT 12
MM10 HUMAN STANDARD; PRT; 476 AA.
ID P09238:
AC 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Stromelysin-2 precursor (EC 3.4.24.22) (Matrix metalloproteinase-10)
DE (MMP-10) (Transin-2) (SL-2).
GN MMP10 OR STMW2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88339885; PubMed=2844164;
RA Muller D., Quantin B., Gesnel M.-C., Milon-Collard R., Abecassis J.,
RA Breathnach R.,
RT "The collagenase gene family in humans consists of at least four
RT members."
RL Biochem. J. 253:187-192(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Ovary;
RC MEDLINE=23388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Scheffer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,

```

```

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrino P., Prange C.,
RA Rana S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.O., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: CAN DEGRADE FIBROBLAST GELATIN OF TYPE I, III, IV,
CC -!- AND V. WEAKLY COLLAGENS III, IV, AND V. ACTIVATES PROCOLLAGENASE.
CC -!- CATALYTIC ACTIVITY: Similar to stromelysin 1, but acts on
CC -!- collagen types III, IV and V is weak.
CC -!- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: X07820; CA30679.1; -.
DR EMBL: BC002591; AA02591.1; -.
DR PIR: A28816; KCHUS2.
DR HSSP: P08254; 1SLM.
DR GenPe: M10.006; -.
DR MIM: 185260; -.
DR GO: GO:0005578; C:extracellular matrix; TAS.
DR GO: GO:0005615; C:extracellular space; TAS.
DR GO: GO:0008270; F:zinc ion binding activity; TAS.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001818; Matrxin.
DR InterPro: IPR006026; N2A MTpeptidase.
DR InterPro: IPR006025; Zn MTpeptidase.
DR Pfam: PF00045; hemopexin; 4.
DR Pfam: PF00413; peptidase_M10; 1.
DR Pfam: PF03933; peptidase_M10_N; 1.
DR PRINTS: PR00138; MATRXIN.
DR SMART: SMO0120; HX; 4.
DR SMART: SMO0235; ZNMG; 1.
DR PROSITE: PS00024; HEMOPEXIN; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; 1.
KW Hydroxylase; Metalloprotease; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal.
FT SIGNAL 1 17
FT PROPEP 18 98
FT CHAIN 99 476
FT DOMAIN 286 476
FT SITE 91 91
FT METAL 217 217
FT ACT SITE 218 218
FT METAL 221 221
FT METAL 227 227
FT DISULFID 289 476
SQ SEQUENCE 476 AA; 54151 MW; 516DCDDFEF92A0D6 CRC64;

Query Match 8.7%; Score 189.5; DB 1; Length 476;
Beet Local Similarity 28.4%; Pred. No. 1e-08;
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

OY 25 PPTSAHGVABGETPD-PDVTGRCSDDGWSFDA-TTLDNGTMLPFKGEFVKSHKMDRE 83

```

```

CC  entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Y13185; CAJ73641.1; -.
DR  PIR; J06505; J06505.
DR  HSSP; P08254; ISLM.
DR  MEROPS; M10.006; -.
DR  MGD; MGI:97007; Mmp10.
DR  InterPro; IPR001843; Fragilysin.
DR  InterPro; IPR000585; Hemopexin.
DR  InterPro; IPR001818; Matrxin.
DR  InterPro; IPR006026; Nzn_MTPeptide.
DR  InterPro; IPR006025; Zn_MTPeptide.
DR  Pfam; PF00045; hemopexin; 4.
DR  Pfam; PF00413; peptidase_M10; 1.
DR  Pfam; PF03933; peptidase_M10_N; 1.
DR  PRINTS; PR00138; MATRXIN.
DR  SMART; SM00120; HX; 3.
DR  SMART; SM00235; ZmC; 1.
DR  PROSITE; PS00024; HEMOPEXIN; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR  Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium;
DR  Collagen degradation; Extracellular matrix; Signal.
FT  SIGNAL 1 17
FT  PROPEP 1 99
FT  CHAIN 100 476
FT  DOMAIN 286 476
FT  SITE 92 92
FT  METAL 218 218
FT  ACT SITE 219 219
FT  METAL 222 222
FT  METAL 228 228
FT  DISULFID 289 476
SQ  SEQUENCE 476 AA; 53911 MW; 2EB1CC41468F0AC6 CRC64;

Query Match 8.94; Score 195.5; DB 1; Length 476;
Best Local Similarity 26.34; Pred. No. 3.2e-09;
Matches 65; Conservative 26; Mismatches 85; Indels 71; Gaps 10;

QY 23 ATPLPPTSAHGNVAEGTKEPDVTERKCSQGSFDA-TLLDNGTMLFFKGEFVMSHKV 81
DB 271 ATVPVLS-----VSPRPETPDCKDPALEFSDSTL--RGEVLFVFDRIKFWRSW 319
QY 82 DRE---LISERKKNPSPVDAAFRQGN--SVFLIGDKVWVYPRPKKEKGYPKLID- 134
DB 320 NPEPEPHLISAFWPTLPDIDAAV-EAHNTDSVLIFFKGSQFMVAVRGNEVQAGYPKGIHTL 378
QY 135 EFGGISPLDAAVECHRGEGCAEGVLFQGHGRNGCHGNSHGHGEVWRCSPHVLVLSA 194
DB 379 GFPPYTKKIDAAV-----FEKEKK----- 398
QY 195 LTSDNMGATYAFSGTHYRLDTSR---DGNHSPIAHQWQPSADVDAFSEWEKLYLVQ 251
DB 399 -----TYFFVGDKVMRDETRHWWDKGRPRQITDDPFGISRPVDVALVHGEGFFFR 450
QY 252 GTQVYVF 258
DB 451 GSSQFER 457

RESULT 11
MM03_HORSE STANDARD; PRT; 477 AA.
AC Q28397;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3)
DE (MMP-3).
GN MMP3.
OS Equus caballus (Horse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OC NCBI_Taxid=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=99074117; PubMed=9858406;
RA Richardson D.W., Dodge G.R.;
RT "Molecular characteristics of equine stromelysin and the tissue
RT inhibitor of metalloproteinase 1."
RL Am. J. Vet. Res. 59:1557-1562(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=98104320; PubMed=9442239;
RA Ballman C.E., Nixon A.J.;
RT "Molecular cloning and cartilage gene expression of equine stromelysin
RT 1 (matrix metalloproteinase 3).";
RL Am. J. Vet. Res. 59:30-36(1998).
RN [3]
RP 3D-STRUCTURE MODELING.
RA Mallena S.C., Sharma J.A.R.P.;
RT "Theoretical model of horse stromelysin."
RL Submitted (MAR-2002) to the PDB data bank.
CC -1- FUNCTION: CAN DEGRADE FIBRONECTIN, LAMININ, GELATINS OF TYPE I,
CC III, IV, AND V; COLLAGENS III, IV, X, AND IX, AND CARTILAGE
CC PROTEOLYCN'S. ACTIVATES PROCOLLAGENASE.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage where P1', P2' and P3'
CC are hydrophobic residues.
CC -1- COPACITOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC -1- SIMILARITY: Contains 1 hemopexin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; U62529; AAB05774.1; -.
DR  PDB; 1L91; 03-APR-02.
DR  MEROPS; M10.005; -.
DR  InterPro; IPR000585; Hemopexin.
DR  InterPro; IPR001818; Matrxin.
DR  InterPro; IPR006026; Nzn_MTPeptide.
DR  InterPro; IPR006025; Zn_MTPeptide.
DR  Pfam; PF00045; hemopexin; 4.
DR  Pfam; PF00413; peptidase_M10; 1.
DR  Pfam; PF03933; peptidase_M10_N; 1.
DR  PRINTS; PR00138; MATRXIN.
DR  SMART; SM00120; HX; 4.
DR  SMART; SM00235; ZmC; 1.
DR  PROSITE; PS00024; HEMOPEXIN; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR  Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
DR  Collagen degradation; Extracellular matrix; Signal; 3D-structure.
FT  SIGNAL 1 17
FT  PROPEP 1 99
FT  CHAIN 100 477
FT  DOMAIN 287 477
FT  SITE 92 92
FT  METAL 218 218
FT  ACT SITE 219 219
FT  METAL 222 222
FT  METAL 228 228
FT  CARBOHYD 120 120
FT  DISULFID 290 477
FT  CONFLICT 346 346
FT  TURN 104 105
FT  STRAND 113 118

```

```

Db 337 T--HDAVA-QIRGAFFFKGYFWRLLR-DRHLVSLDPAQWHRFRGIPHLSDVAVY 392
OY 102 -ROGNSVFLKGDVKVWVYPPREKKEGYPKLLDPEFGIPSLDAVAVCHGECAEVL 160
Db 393 ERTSDHKLVFPRKGRDYWFKNVNEEGYPRVSD-FSLPPGSDIDA----- 437
OY 161 FFOGHGHNCTGNGNSTHGEYMRCSPHLVLSALTSNNGATYAFSGTHYRLDT-SRD 219
Db 438 FSWAHNDR-----TYFFKQQLVWRDYDHTRH 463
OY 220 GWHSMPT-AHQWPGPSAVDAAFSWEE-KLVLQSTQYVYPLTKGTYLVSGYPRLEKE 277
Db 464 MDGPYPAPSPMRGVPTLDDAMRWSDGASYFFRG-QEYWKVLDELVAQVYPOSTARD 522
OY 278 -----VGTPHGIILSDVAA 292
Db 523 WLVCGDSDQSDGSVAAGVDA 542

RESULT 9
MM12_RABIT
ID MM12_RABIT STANDARD; PRT; 464 AA.
AC P79227;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrophage metalloelastase precursor (EC 3.4.24.65) (MME) (Matrix
  metalloproteinase-12) (MMP-12).
GN MMP12.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand;
RA Hou P.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN TISSUE INJURY AND REMODELING. HAS
  SIGNIFICANT ELASTOLYTIC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of soluble and insoluble elastin.
  Specific cleavages are also produced at 14-Ala-[Leu-15 and 16-
  Tyr-[Leu-17 in the B chain of insulin.
CC -1- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC -1- SIMILARITY: Contains 1 hemopexin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U88652; AAB46993.1; -.
DR HSSP; P03956; ICGL.
DR MEROPS; M10.009; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrxin.
DR InterPro; IPR006026; N2a MTpeptide.
DR InterPro; IPR006025; Zn MTpeptide.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; peptidase_M10_1; 1.
DR Pfam; PF03933; peptidase_M10_N; 1.
DR PRINTS; PR00138; MATRXIN.
DR SMART; SM00120; HX; 4.
DR SMART; SM00235; Zmc; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW Hydroxylase; Metalloproteinase; Glycoprotein; Zinc; Zymogen; Calcium;

```

```

KW Extracellular matrix; Signal.
FT SIGNAL 1 17 PROBABLE.
FT PROPE 18 100 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 101 464 MACROPHAGE METALLOELASTASE.
FT DOMAIN 274 464 HEMOPEXIN-LIKE.
FT SITE 87 87 CYSTEINE SWITCH (BY SIMILARITY).
FT METAL 213 213 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 214 214 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 217 217 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 223 223 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 69 69 N-LINKED (GLCNAC...) (POTENTIAL).
FT DISULFID 277 464 BY SIMILARITY.
SQ SEQUENCE 464 AA; 53584 MW; BC8398FD97A15E CRC64;

Query Match 9.0%; Score 196; DB 1; Length 464;
Best Local Similarity 27.4%; Pred. No. 2.8e-09;
Matches 64; Conservative 19; Mismatches 91; Indels 60; Gaps 8;

OY 30 SAHGNVABGETKPPDYTE--RCSDGWSFDATLIDNGTMLFPKGEFVW-----KSHKMDR 83
Db 255 SLVCGPEHQHMPKRDNPDEPTACDHNLKFDVTTVGN-KIFFPDSFWMKIPSSSTTSV 313
OY 84 ELISERKKNPSPYDAAFROG-HNSVFLKGDKWTYPPREKKEGYPKLLD-EFGIGPS 141
Db 314 RLISLWPTLPSGIEAAVEIGDRHQVFLFKGDKFWLISHRLQPNYKSIHSLGFPFVK 373
OY 142 PLDAVAVCHGECAEVLFFQGHGHNCTGNGNSTHGEYMRCSPHLVLSALTSNNG 201
Db 374 KIDPAV-----FNSPL-----R 385
OY 202 ATYAFSGTHYRLDTSR---DGHSMPIAHQWPGPSAVDAAFSEWEKLYLVQG 252
Db 386 KTFEFDVNLVWRDYERREWDAGYPKLITKHPEIGKIDAVFQGYVYFPG 439

RESULT 10
MM10_MOUSE
ID MM10_MOUSE STANDARD; PRT; 476 AA.
AC O55123;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Stromelysin-2 precursor (EC 3.4.24.22) (Matrix metalloproteinase-10)
  (MMP-10) (Transin-2) (SL-2).
GN MMP10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA MEDLINE=98087420; Pubmed=9427548;
RA Madlener M., Werner S.;
RT "cDNA cloning and expression of the gene encoding murine stromelysin-2
  (MMP-10).";
RT Gene 202:75-81(1997).
CC -1- FUNCTION: CAN DEGRADE FIBRONECTIN, GELATINS OF TYPE I, III, IV,
  AND V, WEAKLY COLLAGENS III, IV, AND V. ACTIVATES PROCOLLAGENASE.
CC -1- CATALYTIC ACTIVITY: Similar to stromelysin 1, but action on
  collagen types III, IV and V is weak.
CC -1- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE. WEAK LEVELS IN
  HEART AND LUNG.
CC -1- INDUCTION: By wounding.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC -1- SIMILARITY: Contains 1 hemopexin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial

```

RESULT 8  
MM17\_HUMAN STANDARD; PRT; 606 AA.  
ID 09UL29; Q14850;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Matrix metalloproteinase-17 precursor (EC 3.4.24.-) (MMP-17)  
DE (Membrane-type matrix metalloproteinase 4) (MT-MMP 4) (Membrane-type-4  
DE matrix metalloproteinase) (MT4-MMP).  
GN MMP17 OR MT4MMP.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM LONG).  
RC TISSUE=Monocytic leukemia;  
RX MEDLINE=99402951; PubMed=10471807;  
RA Kajita M., Kinoh H., Ito N., Takamura A., Itoh Y., Okada A., Sato H.,  
RA Seiki M.;  
RT "Human membrane type-4 matrix metalloproteinase (MT4-MMP) is encoded  
RT by a novel major transcript: isolation of complementary DNA clones  
RT for human and mouse mt4-mmp transcripts.";  
RL FEBS Lett. 457:353-356(1999).  
RN [2]  
RP REVISIONS TO 44; 47; 205-210; 224 AND 228.  
RA Seiki M.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=96234364; PubMed=8640782;  
RA Puente X.S., Pendas A.M., Llano E., Velasco G., Lopez-Otin C.;  
RT "Molecular cloning of a novel membrane-type matrix metalloproteinase  
RT from a human breast carcinoma.";  
RL Cancer Res. 56:944-949(1996).  
RN [4]  
RP SEQUENCE OF 129-302 FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=20020281; PubMed=10551873;  
RA Wang Y., Johnson A.R., Ye Q.-Z., Dyer R.D.;  
RT "Catalytic activities and substrate specificity of the human membrane  
RT type 4 matrix metalloproteinase catalytic domain.";  
RL J. Biol. Chem. 274:33043-33049(1999).  
RN [5]  
RP GPI-ANCHOR.  
RX MEDLINE=20036570; PubMed=10567400;  
RA Itoh Y., Kajita M., Kinoh H., Mori H., Okada A., Seiki M.;  
RT "Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a  
RT glycosylphosphatidylinositol-anchored proteinase.";  
RL J. Biol. Chem. 274:34260-34266(1999).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=20008793; PubMed=10543448;  
RA Kolbenbrock H., Essers L., Ubrich N., Will H.;  
RT "Biochemical characterization of the catalytic domain of membrane-type  
RT 4 matrix metalloproteinase.";  
RL Biol. Chem. 380:1103-1108(1999).  
CC -1- FUNCTION: Endopeptidase that degrades various components of the  
CC extracellular matrix, such as fibrin. May be involved in the  
CC activation of membrane-bound precursors of growth factors or  
CC inflammatory mediators, such as tumor necrosis factor-alpha. May  
CC also be involved in tumoral process. Not obvious if able to  
CC proteolytically activates progelatinase A. Does not hydrolyze  
CC collagen types I, II, III, IV and V, gelatin, fibronectin,  
CC laminin, decorin nor alpha1-antitrypsin.  
CC -1- CATALYTIC ACTIVITY: Cleaves pro-TNF-alpha at the 74-Ala-Gln-75  
CC site.  
CC -1- COFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;

CC Name=Long;  
CC IsoId=09UL29-1; Sequence=Displayed;  
CC Name=Short; Synonyms=Puente;  
CC IsoId=09UL29-2; Sequence=VSP 005456;  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LEUKOCYTES, COLON, OVARY  
CC TESTIS AND BREAST CANCER. EXPRESSED ALSO IN MANY TRANSFORMED AND  
CC NON-TRANSFORMED CELL TYPES.  
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FIRIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.  
CC -1- SIMILARITY: Contains 1 hemopexin-like domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AB021225; BA83707.2; -.  
CC EMBL: X89576; CA61753.1; -.  
CC HSSP: Q02853; IHV5.  
CC MEROPS: M10.017; -.  
CC GeneW: HGNC:7163; MMP17.  
CC MIM: 602285; -.  
CC DR GO: GO:0005887; C: integral to plasma membrane; TAS.  
CC DR GO: GO:0008047; E: enzyme activator activity; TAS.  
CC DR GO: GO:0004222; F: metalloendopeptidase activity; TAS.  
CC DR GO: GO:0008270; F: zinc ion binding activity; TAS.  
CC DR InterPro: IPR000585; Hemopexin.  
CC DR InterPro: IPR001818; Matrxin.  
CC DR InterPro: IPR006026; Nzn\_MTPeptidse.  
CC InterPro: IPR006025; Zn\_MTPeptidse.  
CC Pfam: PF00045; hemopexin; 4.  
CC Pfam: PF00413; Peptidase\_M10; 1.  
CC Pfam: PF03933; Peptidase\_M10\_N; 1.  
CC PRINTS: PR00138; MATRIXIN.  
CC SMART: SM00120; HX; 4.  
CC SMART: SM00235; ZMNC; 1.  
CC DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
CC DR PROSITE: PS00546; CYSTEINE\_SWITCH; FALSE\_NEG.  
CC DR PROSITE: PS00024; HEMOPEXIN; FALSE\_NEG.  
CC KM Hydroxylase; Metalloprotease; Zinc; Calcium; Zymogen; Signal;  
CC KX GPI-anchor; Extracellular matrix; Alternative splicing.  
CC FT SIGNAL 1 38  
CC FT PROPEP 39 128  
CC FT CHAIN 129 568  
CC FT PROPEP 569 606  
CC FT DOMAIN 332 529  
CC FT SITE 113 113  
CC FT METAL 251 251  
CC FT ACT\_SITE 252 252  
CC FT METAL 255 255  
CC FT METAL 261 261  
CC FT DISULFID 335 526  
CC FT LIPID 568 568  
CC FT MAIN 125 128  
CC FT CARBOHYD 140 140  
CC FT CARBOHYD 321 321  
CC FT VARSPLIC 1 87  
CC FT  
CC SQ SEQUENCE 606 AA; 67006 MW; P5D51445118F4DF5 CRC64;  
Query Match 9.1%; Score 199; DB 1; Length 606;  
Best Local Similarity 25.0%; Pred. No. 2, 1e-09;  
Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;  
QY 5 LGAPVALG-----WELCWSLAATPL-----PPTSAHGNGVAGETKPPDPYERCS 51  
DB 277 VGPPLRYGYPDYDKRVWOLYGVRESVSPPTAPBEPPLLPDPDSSAPPKRDVPHKCS 336  
QY 52 DGWSPDATTLDNGTALPFKSGFVWVKSHKMDRELISER-----WKNFP---SPVDAAF 101

```

QY 134 DEFGIPSPDLAAVECHRGCEQAEGLFPOGHGHRNGTGHGNSHTHGEYWRSCPHLVLS 193
DB 427 D-FSLPFGIGDA-----VFMAHNDR-----446
QY 194 ALTSNMHATYAFSGTHYWRD--TSR--DGKHS-WPIAHQWPGQPSAVDAAFSWEK-KL 247
DB 447 -----TYFFKQQLWRYDDHTRMDPGYPAGQGP--WRGVPSMLDAMRWSDGAS 494
QY 248 YLVQGTQYVFLTKGTYLVSGYPRLEKE--VGTPHGILLSDVAFAICGSSRLHIM 304
DB 495 YFFRG-QGYWKLVDLELAARPGYPOSTARDMLVCGEP-----LADADVGGPO-----542
QY 305 AGRLMLDLKSGAQATWTELPWPEKXVDG-ALCMKSLGPNSCSANGPGLYLHGPMLY 363
DB 543 -GR-----SGAQ-----DGLAVCS-----555
QY 364 CYSDVEKNAKALPQPNVTSL 387
DB 556 CTSDAHL-----ALPSLLLTPL 575

RESULT 7
MM12_HUMAN
ID MM12_HUMAN STANDARD; PRT; 470 AA.
AC P39900;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Macrophage metalloelastase precursor (EC 3.4.24.65) (HME) (Matrix
DE metalloproteinase-12) (MMP-12) (Macrophage elastase) (ME).
GN MMP12 OR HME.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Alveolar macrophage;
RX MEDLINE=94043200; PubMed=8226919;
RA Shapiro S.D., Kobayashi D.K., Ley T.J.;
RT "Cloning and characterization of a unique elastolytic
RT metalloproteinase produced by human alveolar macrophages.";
RL J. Biol. Chem. 268:23824-23829 (1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97277370; PubMed=9115292;
RA Gronski T.J., Jr., Martin R.L., Kobayashi D.K., Walsh B.C.,
RA Holman M.C., Huber M., Van Wart H.G., Shapiro S.D.;
RT "Hydrolysis of a broad spectrum of extracellular matrix proteins by
RT human macrophage elastase.";
RL J. Biol. Chem. 272:12189-12194 (1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 100-280.
RX MEDLINE=21460859; PubMed=11575929;
RA Nar H., Werle K., Bauer M.M.T., Dollinger H., Jung B.;
RT "Crystal structure of human macrophage elastase (MMP-12) in complex
RT with a hydroxamic acid inhibitor.";
RL J. Mol. Biol. 312:743-751 (2001).
CC -1- FUNCTION: May be involved in tissue injury and remodeling. Has
CC significant elastolytic activity. Can accept large and small amino
CC acids at the PL' site, but has a preference for leucine. Aromatic
CC or hydrophobic residues are preferred at the PL site, with small
CC hydrophobic residues (preferably alanine) occupying P3.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of soluble and insoluble elastin.
CC Specific cleavages are also produced at 14-Ala-[Leu-15 and 16-
CC Tyr-1]-Leu-17 in the B chain of insulin.
CC -1- COFACTOR: Requires calcium and zinc for activity.
CC -1- TISSUE SPECIFICITY: FOUND IN ALVEOLAR MACROPHAGES BUT NOT IN
CC PERIPHERAL BLOOD MONOCYTES.
CC -1- INDUCTION: BY EXPOSURE TO LYPOPOLYSACCHARIDE. INHIBITED BY
CC DEXAMETHASONE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.

```

```

CC -1- SIMILARITY: Contains 1 hemopexin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DB EMBL, L23808; AA58658.1; ALT_SEQ.
DB PIR, A49499; A49499.
DB PDB, 1J1Z; 03-JUL-02.
DB PDB, 1JX3; 28-SEP-01.
DB MEROPS, M10.009; -.
DB GeneW, HGNC:7156; MMP12.
DR MIM, 601046; .
DR GO, GO:0004234; F:macrophage elastase activity; TAS.
DR GO, GO:0008270; F:zinc ion binding activity; TAS.
DR GO, GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro, IPR00585; Hemopexin.
DR InterPro, IPR001818; Matrixin.
DR InterPro, IPR006026; NZn_MTPeptide.
DR InterPro, IPR005025; Zn_MTPeptide.
DR Pfam, PF00045; hemopexin; 4.
DR Pfam, PF00413; peptidase M10; 1.
DR Pfam, PF03933; peptidase M10_N; 1.
DR PRINTS, PR00138; MATRIXIN.
DR SMART, SM00120; HX; 4.
DR SMART, SM00235; ZnMc; 1.
DR PROSITE, PS00024; HEMOPEXIN; 1.
DR PROSITE, PS00142; ZINC_PROTEASE; 1.
DR PROSITE, PS00546; CYSTEINE_SWITCH; 1.
KW Hydroxylase; Metalloproteinase; Glycoprotein; Zinc; Zymogen; Calcium;
KW Extracellular matrix; Signal; 3D-structure.
FT SIGNAL 1 16
FT PROPEP 17 105
FT CHAIN 106 470
FT DOMAIN 279 470
FT SITE 92 92
FT METAL 218 218
FT ACT_SITE 219 219
FT METAL 222 222
FT METAL 228 228
FT CARBOHYD 20 20
FT CARBOHYD 285 285
FT DISULFID 282 470
SQ SEQUENCE 470 AA; 54001 MW; 8C745E8A8C0EA216 CRC64;
Query Match 9.3%; Score 202.5; DB 1; Length 470;
Best Local Similarity 27.3%; Pred. No. 7; De-10;
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;
QY 30 SAHGNVAEGETKDDPDVTE--RCSDGNSFDATLLDNGTMLFPKGEFVW-----KSHKMDR 83
DB 260 SLVDDPKENQRLPNPDSBPCDNLSPDAVTVGN-KLFFEXDRFMLKVSBRPTSV 318
QY 84 ELISERKYNRPSPDAFR-QGHSVFLIGDKWYVPRPEKKEGYKLLQD-EFPPIPS 141
DB 319 NLISLWPTLPSSGIEAAYEIAERQVFLFDQDKYMLISNLRPEBNYPKSIHSFGFPNVK 378
QY 142 PLDAVAECHRGCEQAEGLFPOGHGHRNGTGHGNSHTHGEYWRSCPHLVLSALTSNMHG 201
DB 379 KIDAAV-----FNRFPR-----391
QY 202 ATYAFSGTHYWRDTSRD---GHSWPIAHQWPGQPSAVDAAFSWEK-LYLVQGTQVY 256
DB 392 -TYFFVQNWRYDDHTRMDPGYPKLTGNPGSIGRK-IDAVFYSKNKYYFFQGSNGF 449
QY 257 VF 258
DB 450 EY 451

```

QY 119 YPEEKEKGYPKLLQDEFPPIPSPLDAVECHRGCOAGVLFPOG----- 164  
 DB 121 YPRE-KEKENPRELQGEFPVPSPLAAVCHRGCECODEVLFQGHMFMWSTTKTK 179  
 QY 165 -----HGH-----RNGTGH 173  
 DB 180 ERLMPAVNCSSAMRWISRYCYCRGNQFLRFPDVTGHVDPKYPDRVDYFMSCGRGHAH 239  
 QY 174 GNSHTHGPEYMRCSPLHVSALTSNDHGATYASGTHYMLDTRSDGMSWPIAHQMPQ 233  
 DB 240 RIAHTHGDD--RCSPLVLTALLSDNHGATYARGHYHRLDTSRDGMSWPIAHQMSHG 297  
 QY 234 PSAVDAFMSBEKLYLVQGTQVYVFLTKGQYTLVSGYPKLEKXVGTPHGILDSVDAF 293  
 DB 298 PSAVDAFMSWDDKLYLQGTQVYIFLTGAGYTLVDVYPKQLEKELGSPHGISLDAVDATF 357  
 QY 294 ICGSSRLHIMAGRLMLDLKSGAQTVELPWPHEKDGALCMKSLGPNCSANGRC 353  
 DB 358 VCGSTRHLHVMARKLWMLDLKSGAQPWTELPWHEKDAALCTEKSLGPNCSASGLG 417  
 QY 354 LYLHGPNIYCSGDEKLNAAKALPOPQNTSLG 389  
 DB 418 LYLHGPNIYCYKDEKLVSAKALPOPQNTSLG 453

RESULT 6  
 MM17\_MOUSE STANDARD; PRT; 578 AA.  
 AC 09R053/  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Matrix metalloproteinase-17 precursor (EC 3.4.24.-) (MMP-17)  
 DE (Membrane-type matrix metalloproteinase 4) (MT-MMP 4) (Membrane-type-4  
 matrix metalloproteinase) (MT4-MMP).  
 GN MMP17 OR MT4MP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic brain;  
 RX MEDLINE=99402951; PubMed=10471807;  
 RA Kajita M., Kinoh H., Ito N., Takamura A., Itoh Y., Okada A., Sato H.,  
 RT "Human membrane type-4 matrix metalloproteinase (MT4-MMP) is encoded  
 by a novel major transcript: isolation of complementary DNA clones for  
 human and mouse mt4-mmp transcripts.";  
 RL FEBS Lett. 457:353-356(1999).  
 RN [2]  
 RP REVISIONS TO C-TERMINUS.  
 RA Seiki M.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=20261520; PubMed=10799478;  
 RA English W.R., Puente X.S., Freije J.M.P., Knaeuper V., Amour A.,  
 RA Merryweather A., Lopez-Otin C., Murphy G.,  
 RT "Membrane type 4 matrix metalloproteinase (MMP17) has tumor necrosis  
 factor-alpha convertase activity but does not activate pro-MMP2.";  
 RL J. Biol. Chem. 275:14046-14055(2000).  
 RN [4]  
 RP GPI-ANCHOR, AND MUTAGENESIS OF GLU-248.  
 RX MEDLINE=20036570; PubMed=10567400;  
 RA Itoh Y., Kajita M., Kinoh H., Mori H., Okada A., Seiki M.;  
 RT "Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a  
 glycosylphosphatidylinositol-anchored proteinase.";  
 RL J. Biol. Chem. 274:34260-34266(1999).  
 CC -!- FUNCTION: Endopeptidase that degrades various components of the  
 CC extracellular matrix, such as fibrin. May be involved in the  
 CC activation of membrane-bound precursors of growth factors or

CC inflammatory mediators, such as tumor necrosis factor-alpha. May  
 CC also be involved in tumoral process. Not obvious if able to  
 CC proteolytically activate progelatinase A. Does not hydrolyze  
 CC collagen types I, II, III, IV and V, gelatin, fibronectin,  
 CC laminin, decorin nor alpha1-antitrypsin.  
 CC -!- CATALYTIC ACTIVITY: Cleaves pro-TNF-alpha at the 74-Ala-I-Gln-75  
 CC site.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit, calcium (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY MONOCYTES AND MACROPHAGES.  
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.  
 CC -!- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; AB021224; BAA82708.2; -.  
 DR EMBL; AJ010731; CAB92315.1; -.  
 DR HSSP; Q02853; 1HV5.  
 DR MEROPS; M10.017; -.  
 DR MGD; MGI:1346076; MMP17.  
 DR InterPro; IPR000385; Hemopexin.  
 DR InterPro; IPR001818; Matrixin.  
 DR InterPro; IPR006026; Nzn\_MTPeptide.  
 DR InterPro; IPR006025; Zn\_MTPeptide.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF00413; Peptidase\_M10; 1.  
 DR Pfam; PF03933; Peptidase\_M10\_N; 1.  
 DR PRINTS; PR00136; MATRIXIN.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00235; Zmnc; 1.  
 DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE\_NEG.  
 DR PROSITE; PS00024; HEMOPEXIN; FALSE\_NEG.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; Zinc; Calcium; Zymogen; Signal;  
 KW GPI-anchor; Extracellular matrix.  
 KM  
 FT SIGNAL 1 39  
 FT PROPEP 125 124  
 FT CHAIN 125 558  
 FT PROPEP 559 578  
 FT DOMAIN 333 530  
 FT SITE 109 109  
 FT METAL 247 247  
 FT ACT SITE 248 248  
 FT METAL 251 251  
 FT METAL 251 251  
 FT METAL 257 257  
 FT DISULFID 336 527  
 FT LIPID 558 558  
 FT CARBOHYD 136 136  
 FT CARBOHYD 322 322  
 FT MUTAGEN 248 248  
 FT CONFLICT 45 45  
 FT CONFLICT 277 277  
 FT SEQUENCE 578 AA; 64319 MW; 09AA933B6A02178 CRC64;

Query Match 9.3%; Score 203; DB 1; Length 578;  
 Best Local Similarity 25.0%; Pred. No. 9, 1e-10;  
 Matches 96; Conservative 33; Mismatches 109; Indels 146; Gaps 24;

QY 25 PLPTPSAHGNVABGETKPPDYTERCSDSGMSFDATTLDDNGTMLPFKGFVWKSHPKMBE 84  
 DB 317 PEPPNN-----RSSRPCKDVPHRCTA--HFDVAVA-QRGEAFEFKGYFWRLTR-DBH 366  
 QY 85 LLSER-----WKQPF--SPVDAF-RQGNVSVLIGDKQWVYPPPEKKEKGYPKLLQ 133  
 DB 367 LVSLQPAQMRHFWRGPLHLDSVDVAVERTSDHKIVFFKGRDVRWPKDNVVEGYPKPV 426

KM Glycoprotein; Heme; Plasma; Repeat; Transport; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 460  
 FT DOMAIN 56 93  
 FT DOMAIN 97 140  
 FT DOMAIN 187 230  
 FT DOMAIN 261 304  
 FT DOMAIN 306 349  
 FT METAL 79 79  
 FT METAL 149 149  
 FT DISULFID 50 230  
 FT DISULFID 148 153  
 FT DISULFID 187 199  
 FT DISULFID 255 458  
 FT DISULFID 364 406  
 FT DISULFID 416 433  
 FT CARBOHYD 38 38  
 FT CARBOHYD 64 64  
 FT CARBOHYD 186 186  
 FT CARBOHYD 240 240  
 FT CARBOHYD 246 246  
 FT CARBOHYD 419 419  
 FT CONFLICT 30 30  
 FT CONFLICT 233 233  
 FT CONFLICT 453 453  
 SQ SEQUENCE 460 AA; 51340 MW; 363BAC7520D1E39 CRC64;  
 Query Match 70.7%; Score 1545.5; DB 1; Length 460;  
 Best Local Similarity 62.9%; Pred. No. 5e-125;  
 Matches 292; Conservative 38; Mismatches 55; Indels 79; Gaps 6;  
 QY 1 MARVLGAPVALGLMSLWLSLAIAATPLPTSAHGNVAGEE--TKDDPVTETRCSDGWSFDA 58  
 DB 1 MARVAVALNIIIVLGLCSLAIVASPLP--TAHGRAVAEVENGTKDDPVTETRCSDGWSFDA 58  
 QY 59 TITLDNGTMTLFFKGEFVWKSFKMDRELISERWKPSPSPVDAAFROGANSVFLIKGDKWV 118  
 DB 59 ATMDHNGTMTLFFKGEFVWKSFKMDRELISERWKPSPSPVDAAFR-GPDVFLIKGDKWV 117  
 QY 119 YPEKKEKGVYKLLQDEFPFGIPSPUDAAVECHRGCEQAGVLPFGQH----- 165  
 DB 118 YPEKKEKGVYKLLQDEFPFGIPSPUDAAVECHRGCEQAGVLPFGQNKMKWMDPATRTQK 177  
 QY 166 -----GH-- 167  
 DB 178 ERSWSTVGNCTALRWLERYYCFOGNKFLRFNPVTGEVPRPYPLDARDYFVSCRGHGR 237  
 QY 168 -RNGTGAGNSTHHPETWRCSPHLVLSALTSNDHGAITYAEGTTHWRIDTSDGWSHWP 226  
 DB 238 PRNGTAAHGNSTH--PMHSRCSPPDGLTALSDHAGATYAFGSHYWRIDSSRDGWSHWP 295  
 QY 227 AHOWPGPSAVDAAFSWEKLYLVQGTQVYVFLTKGGTIVSGYPKRLKEKVGTPHGIIL 286  
 DB 296 AHHPQGPSTYDAFSDWDKYLIGQTVYVFLTKGGNNLVSGYPKRLKEKLSPPGISL 355  
 QY 287 DSVDAFICPSSSLHIMAGRLWMLDLKSGAQTWELPWPHEKVGALCMESLIGNS 346  
 DB 356 ETDAARSCPSSSLYVSSGRMLWMLDLKSGAQTWELPWPHEKVGALCMESLIGNT 415  
 QY 347 CSANGPGLYLHGNLYCYSDVEKLNAAKALPQONTSLIGCT 390  
 DB 416 CSSNGSLYFLHGNLYCYSDIDKLNAAKSLPQOKVNSLIGCS 459  
 RESULT 5  
 HEMO\_PIG STANDARD; PRT; 459 AA.  
 AC P50828;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hemopexin precursor (Hyaluronidase) (EC 3.2.1.35).  
 GN HPX.

OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-48 AND 368-388.  
 RC TISSUE=Liver;  
 RX MEDLINE=95096047; PubMed=7798203;  
 RA Zhu L., Hope T.J., Hall J., Davies A., Stern M.,  
 RA Mueller-Eberhard U., Stern R., Parslow T.G.;  
 RT "Molecular cloning of a mammalian hyaluronidase reveals identity with  
 hemopexin, a serum heme-binding protein."  
 RL J. Biol. Chem. 269:32092-32097(1994).  
 CC -1- FUNCTION: Binds heme and transports it to the liver for breakdown  
 and iron recovery, after which the free hemopexin returns to the  
 circulation.  
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1->4-linkages between N-  
 acetyl-beta-D-glucosamine and D-glucuronate residues in  
 hyaluronate.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U14751; AAC48457.1; -.  
 DR PIR; A55486; A55486.  
 DR HSSP; P20058; IHXN.  
 DR InterPro; IPR00585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 5.  
 DR SMART; SM00120; HX; 5.  
 DR PROSITE; PS00024; HEMOPEXIN; 2.  
 KW Glycoprotein; Heme; Plasma; Repeat; Transport; Signal; Hydrolase;  
 KW Glycosidase.  
 FT SIGNAL 1 28  
 FT CHAIN 29 459  
 FT DOMAIN 58 95  
 FT DOMAIN 99 142  
 FT DOMAIN 189 232  
 FT DOMAIN 256 299  
 FT DOMAIN 301 344  
 FT METAL 81 81  
 FT METAL 151 151  
 FT METAL 151 151  
 FT DISULFID 52 232  
 FT DISULFID 150 155  
 FT DISULFID 189 201  
 FT DISULFID 250 453  
 FT DISULFID 359 401  
 FT DISULFID 411 428  
 FT CARBOHYD 40 40  
 FT CARBOHYD 82 82  
 FT CARBOHYD 188 188  
 FT CARBOHYD 241 241  
 SQ SEQUENCE 459 AA; 51305 MW; DB06B44C29789CF CRC64;  
 Query Match 70.7%; Score 1544; DB 1; Length 459;  
 Best Local Similarity 63.4%; Pred. No. 6.8e-125;  
 Matches 289; Conservative 33; Mismatches 64; Indels 70; Gaps 5;  
 QY 1 MARVLGAPVALGLMSLWLSLAIAATPLPTSA--HGNVAGEETKDDPVTETRCSDGWSFDA 58  
 DB 1 MARVAVALNIIIVLGLCSLAIVASPLP--TAHGRAVAEVENGTKDDPVTETRCSDGWSFDA 60  
 QY 59 TITLDNGTMTLFFKGEFVWKSFKMDRELISERWKPSPSPVDAAFROGANSVFLIKGDKWV 118  
 DB 61 TITLDHNGTMTLFFKGEFVWKSFKMDRELISERWKPSPSPVDAAFRRGHDRVFLIGDKWV 120



```

Db 356 DTIDAAFCSCGSKLVTYSRRLMWLIDKSGAQTAAELBMPHEKVDGALCEKSLGPPYS 415
QY 347 CSANGPELYIHGPNLYCYSVDEKLNAAKALPOPONVTSILGCT 390
Db 416 CSSNGPNLPFIHGPNLYCYSIDKLNAAKSLPOPONKNSITLGS 459

RESULT 4
HEMO MOUSE STANDARD; PRT; 460 AA.
AC 091X72; P97824; OSWUP0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hemopexin precursor.
DE HPX OR HPXN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver:
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.I., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carrincci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield J.S.N., Krzywinski M.I., Skalek U., Smalhe D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE OF 7460 FROM N.A.
RC TISSUE=Liver:
RA Koepsel R.R., Rohrbach D.H., Brekhelner B.B.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN
CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE
CC CIRCULATION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.
CC -1- SIMILARITY: Contains 5 hemopexin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration bet-
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; BC011246; AAH11246.1; -
DR EMBL; BC019901; AAH19901.1; -
DR EMBL; U89889; AAB49490.1; -.
DR MGD; MGI:105112; Hpxn.
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF000045; hemopexin; 5.
DR SMART; SM00120; HX; 5.
DR PROSITE; PS00024; HEMOPEXIN; 1.

```

```

CC      or send an email to: license@ieb-5tb.ch).
-----
DR      EMBL, X16429, CAJ34452.1;  .
DR      PDB; 1HXN; 15-OCT-95.
DR      PDB; 1QS; 03-FEB-00.
DR      InterPro; IPR000585; Hemopexin.
DR      Pfam; PF00045; hemopexin; 5.
DR      SMART; SMO0120; HX; 5.
DR      PROSITE; PS00024; HEMOPEXIN; 1.
KM      Glycoprotein; Heme; Plasma; Repeat; Transport; Signal; 3D-structure.
FT      SIGNAL 1 25
FT      CHAIN 26 460
FT      DOMAIN 58 95 HEMOPEXIN.
FT      DOMAIN 99 143 HEMOPEXIN-LIKE 1.
FT      DOMAIN 190 233 HEMOPEXIN-LIKE 2.
FT      DOMAIN 261 304 HEMOPEXIN-LIKE 3.
FT      DOMAIN 306 349 HEMOPEXIN-LIKE 4.
FT      METAL 81 81 HEMOPEXIN-LIKE 5.
FT      METAL 152 152 IRON (HEME AXIAL LIGAND) .
FT      DISULFID 52 233 IRON (HEME AXIAL LIGAND) .
FT      DISULFID 151 156 BY SIMILARITY.
FT      DISULFID 190 202 BY SIMILARITY.
FT      DISULFID 255 458 BY SIMILARITY.
FT      DISULFID 364 406 BY SIMILARITY.
FT      DISULFID 416 433 BY SIMILARITY.
FT      CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT      CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT      CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT      CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT      CONFLICT 52 52 C -> W (IN REF. 2) .
FT      HELIX 253 255
FT      TURN 257 258
FT      STRAND 263 266
FT      TURN 268 269
FT      STRAND 272 276
FT      TURN 277 278
FT      STRAND 279 282
FT      TURN 289 290
FT      STRAND 293 294
FT      TURN 295 297
FT      HELIX 298 298
FT      TURN 300 301
FT      STRAND 308 312
FT      TURN 313 314
FT      STRAND 315 320
FT      TURN 321 322
FT      STRAND 323 328
FT      STRAND 334 335
FT      TURN 337 338
FT      STRAND 341 342
FT      HELIX 343 347
FT      STRAND 360 362
FT      TURN 365 366
FT      STRAND 369 374
FT      TURN 375 376
FT      STRAND 377 382
FT      HELIX 383 388
FT      STRAND 392 394
FT      STRAND 403 407
FT      STRAND 422 427
FT      TURN 428 429
FT      STRAND 430 434
FT      HELIX 437 442
FT      STRAND 450 451
FT      HELIX 452 455
FT      TURN 456 457
SO      SEQUENCE 460 AA; 51767 MW; 193B59856DEF64EE CRC64;

Query Match 77.8%; Score 1700.5; DB 1; Length 460;
Best Local Similarity 67.7%; Pred.No. 2.6e-136;
Matches 313; Conservative 26; Mismatches 50; Indels 73; Gaps 3,
1 MARVLGARVALGWSLCSLAIATPLPTSAHGNVAEE--TKPPDYTERCSDGWSFDA 58

```

Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
Qy	59	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
Qy	119	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
Qy	230	232	234	236	238	240	242	244	246	248	250	252	254	256	258	260	262	264	266	268	270	272	274	276	278	280	282	284	286	288	290	292	294	296	298	300	302	304	306	308	310	312	314	316	318	320	322	324	326	328	330	332	334	336	338	340	342	344	346	348	350
Db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	

PN WO200166766-A2.  
 XX 13-SEP-2001.  
 XX 06-MAR-2001; 2001WO-US07167.  
 XX 06-MAR-2000; 2000US-0187196.  
 XX (DARW-) DARWIN MOLECULAR CORP.  
 XX (SCHA/) SCHATZMAN R.  
 PI Fajardo M, Wang K, Smith R, Moss P;  
 XX WPI; 2001-58276/65.  
 DR Novel isolated matrix metalloproteinase-25 nucleic acid molecule and  
 PT proteins encoded by them whose inhibition is useful for modulation of  
 PT hair growth in mammals -  
 XX  
 PS Example 2; Fig 3; 11pp; English.  
 XX The present sequence is human matrix metalloproteinase (MMP)-12 protein  
 CC used in the exemplification of the invention. MMP-25 DNA is located  
 CC on chromosome 11q22. Matrix metalloproteinases are a family of zinc  
 CC dependent endopeptidases that function extracellularly to degrade  
 CC proteins typically found in the extracellular matrix. MMP-25 is expressed  
 CC in skin cells of mammals, particularly in breast cells and hair  
 CC follicles. MMP-25 DNA is useful for identifying a nucleic acid molecule  
 CC encoding all or part of MMP by hybridizing MMP-25 to a nucleic acid  
 CC sample and identifying a sequence that hybridizes in the nucleic  
 CC acid sample. The identification step involves performing polymerase  
 CC chain reaction (PCR) to amplify the hybridizing sequence. MMP-25  
 CC antibody is useful for identifying type 25 MMP. MMP-25 protein  
 CC inhibitors may be used to modulate hair growth and breast cancer in  
 CC a mammal.  
 CC  
 SQ Sequence 470 AA;  
 Query Match 9.3%; Score 202.5; DB 22; Length 470;  
 Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
 Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
 QY 30 SAHGVNAEGRTKDPDVT--RCSGMSPDATLTDNGTMLFKGEFVW---KSHKMDR 83  
 DB 260 SLVGDPEKNQRLPNPNSEPALCDPNLSFDAVTVGN-KIFFEKDFFWLKVESRPTSV 318  
 QY 84 ELISERKKNPSPVDAAFR-QGHSVFLIKGDKVWVYPPKKEKGYPKLQD-EFPGIPS 141  
 DB 319 NLISSLMPTLPSCIEAAYIEARNQVFLFDKDKYMLSNLRPEPNYKSHSGFPPFVK 378  
 QY 142 PLDAVECHRGECQAGVLFPGQGHGRNGTGHGNSTHGPEYKCSPLVLALTSNDHG 201  
 DB 379 KIDAAY-----FNRFYR----- 391  
 QY 202 ATYAFSGTHYRLDTSRD-----GMHSWPIAHQMPQGSAYDAAPSWEEK-LYLVQGTQVY 256  
 DB 392 -TYFPVNDQWYRDERROMDPPGYPKLITKNPQIGPK-IDAVYYSKNKYYFFQGSNOF 449  
 QY 257 VF 258  
 DB 450 EY 451  
 RESULT 7  
 AAB84614  
 ID AAB84614 standard; Protein; 470 AA.  
 XX AAB84614;  
 AC AAB84614;  
 XX 05-SEP-2001 (first entry)  
 DT  
 XX Amino acid sequence of matrix metalloproteinase-12.

KM Growth factor; protein inhibitor; protease; damaged tissue;  
 KM platelet-derived growth factor; PDGF; fibroblast growth factor; FGF;  
 KM connective tissue derived growth factor; CTGF; chrysalin; VEGF;  
 KM keratinocyte-derived growth factor; KGF; epidermal growth factor; EGF;  
 KM transforming growth factor-beta; TGF-beta; matrix metalloproteinase; MMP;  
 KM granulocyte macrophage colony stimulating factor; GM-CSF; uPA;  
 KM vascular endothelial growth factor; urokinase plasminogen activator;  
 KM dermal ulcer; wound.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WO200149309-A2.  
 XX 12-JUL-2001.  
 XX 21-DEC-2000; 2000WO-IB01935.  
 XX 29-DEC-1999; 99GB-0030768.  
 XX (PFI2 ) PRIZER LTD.  
 XX (PFI2 ) PRIZER INC.  
 XX Davies MJ, Huggins JP, McIntosh FS, Occleston NL;  
 PI WPI; 2001-418351/44.  
 DR N-PSDB; AAM28229.  
 XX Composition for the treatment of damaged tissue i.e. chronic wounds and  
 PT dermal ulcers comprises an inhibitor agent i.e. a protease and a growth  
 PT factor -  
 XX  
 PS Disclosure; Page 560; 572pp; English.  
 XX The specification describes a pharmaceutical composition, comprising  
 CC a growth factor, an inhibitor agent, i.e. a protease. The inhibitor  
 CC agent inhibits the action of at least one specific adverse protein,  
 CC i.e. a protease, that is upregulated in a damaged tissue such as a  
 CC wound environment. Growth factors which are included in the composition  
 CC of the invention are platelet-derived growth factor (PDGF), fibroblast  
 CC growth factor (FGF), connective tissue derived growth factor (CTGF),  
 CC keratinocyte-derived growth factor (KGF), transforming growth  
 CC factor-beta (TGF-beta), granulocyte macrophage colony stimulating factor  
 CC (GM-CSF), epidermal growth factor (EGF), vascular endothelial growth  
 CC factor (VEGF), and chrysalin. Inhibitors which are included in the  
 CC composition of the invention include inhibitors of urokinase-type  
 CC plasminogen activator (uPA) and matrix metalloproteinase (MMP). The  
 CC composition is useful for the treatment of chronic damaged tissue, i.e.  
 CC wounds and dermal ulcers. The present sequence represents a human MMP-12,  
 CC and is used to produce the composition of the invention.  
 CC  
 SQ Sequence 470 AA;  
 Query Match 9.3%; Score 202.5; DB 22; Length 470;  
 Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
 Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
 QY 30 SAHGVNAEGRTKDPDVT--RCSGMSPDATLTDNGTMLFKGEFVW---KSHKMDR 83  
 DB 260 SLVGDPEKNQRLPNPNSEPALCDPNLSFDAVTVGN-KIFFEKDFFWLKVESRPTSV 318  
 QY 84 ELISERKKNPSPVDAAFR-QGHSVFLIKGDKVWVYPPKKEKGYPKLQD-EFPGIPS 141  
 DB 319 NLISSLMPTLPSCIEAAYIEARNQVFLFDKDKYMLSNLRPEPNYKSHSGFPPFVK 378  
 QY 142 PLDAVECHRGECQAGVLFPGQGHGRNGTGHGNSTHGPEYKCSPLVLALTSNDHG 201  
 DB 379 KIDAAY-----FNRFYR----- 391  
 QY 202 ATYAFSGTHYRLDTSRD-----GMHSWPIAHQMPQGSAYDAAPSWEEK-LYLVQGTQVY 256  
 DB 392 -TYFPVNDQWYRDERROMDPPGYPKLITKNPQIGPK-IDAVYYSKNKYYFFQGSNOF 449  
 QY 257 VF 258

Db 450 EY 451

RESULT 8  
AAB74595 standard; Protein; 470 AA.ID AAB74595 standard; Protein; 470 AA.  
AC AAB74595;  
XX  
XX 11-JUN-2001 (first entry)  
DT  
XX  
XX Human macrophage metalloelastase HME.  
DE  
XX Human; HME: macrophage metalloelastase; elastolysis; pulmonary emphysema;  
KM inflammatory disease.  
XX  
XX Homo sapiens.  
OS  
XX US6204043-B1.  
PN  
XX 20-MAR-2001.  
PD  
XX  
XX 01-MAR-1995; 95US-0396988.  
PF  
XX  
XX 28-MAY-1993; 93US-0068392.  
PR  
XX  
XX (UNIW ) UNIV WASHINGTON.  
PA  
XX  
XX Shapiro SD;  
PI  
XX  
XX WPI; 2001-289518/30.  
DR  
XX N-PSDB; AAF81624.PT New human macrophage metalloelastase, useful in the normal embryonic  
development, growth, tissue remodeling and tissue repair, particularly  
in studying the pathogenesis of pulmonary emphysema -  
PS Claim 1; Fig 5; 20pp; English.CC The present invention provides the protein and coding sequences of human  
CC macrophage metalloelastase (HME). The protein has elastolytic activity.  
CC HME has a role in tissue remodeling and repair associated with  
CC development and inflammation, and abnormal expression can result in  
CC tumour invasiveness, arthritis and atherosclerosis. It is also thought to  
CC contribute to the pathogenesis of pulmonary emphysema and other  
CC inflammatory destructive diseases. The present sequence is the HME  
CC protein.  
XX  
XX  
SQ Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 22; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10; Indels 63; Gaps 10;  
Matches 66; Conservative 23; Mismatches 90;QY 30 SAHGNVAEGETKDPDVT--RCSDEGSPDATTLDNGTMLFFKGEFVW----KSHKMDR 83  
DB 260 SLVGDPKRENGRLPNPNSEPALCDPNLSFDAVTTVGN-KIFFKDRFFMLKVSERPTSV 318  
QY 84 ELISERKKNFPPSVDAFR-QGHNSVFLIKDKRWVYPPKKEKGYPKLIOD-EFPGIPS 141  
DB 319 NLISLSLWPTLPSSGIEAAYEIEARNOVFLFKDDKYWLISNLRPEPNYPKSIHSFGFNFVK 378  
QY 142 PLDAAYECHRGEQAGVLFQGHGHRNGTGHGNSHTHGPEVWRCSPHLVLSLTSNDHG 201  
DB 379 KIDAAV-----FNPRFYK----- 391QY 202 ATYAFSGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAAFSWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNDQYWRDERQMMDPGPKLITKNFGIGPK-IDAVFYKKNKYYFFQGSNOF 449  
QY 257 VF 258

Db 450 EY 451

RESULT 9  
AAB49982 standard; Protein; 470 AA.ID AAB49982 standard; Protein; 470 AA.  
AC AAB49982;  
XX  
XX 13-MAR-2001 (first entry)  
DT  
XX  
XX Human macrophage metalloelastase.  
DE  
XX Human; macrophage metalloelastase; elastin; matrix degrading enzyme;  
KM emphysema.  
XX  
XX Homo sapiens.  
OS  
XX US6150152-A.  
PN  
XX 21-NOV-2000.  
PD  
XX  
XX 28-MAY-1993; 93US-0068392.  
PF  
XX  
XX 28-MAY-1993; 93US-0068392.  
PR  
XX  
XX (UNIW ) UNIV WASHINGTON.  
PA  
XX  
XX Shapiro SD;  
PI  
XX  
XX WPI; 2001-049090/06.  
DR  
XX N-PSDB; AAC64994.PT Novel human macrophage metalloelastase polynucleotides and polypeptides  
useful for measuring elastin degradation -  
PS Claim 1; Fig 5; 20pp; English.CC The present invention provides the coding and protein sequences for the  
CC human macrophage metalloelastase. This is a matrix degrading  
CC metalloproteinase which has the ability to degrade elastin, and can be  
CC used to measure elastin degradation.  
XX  
XX  
SQ Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 22; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10; Indels 63; Gaps 10;  
Matches 66; Conservative 23; Mismatches 90;QY 30 SAHGNVAEGETKDPDVT--RCSDEGSPDATTLDNGTMLFFKGEFVW----KSHKMDR 83  
DB 260 SLVGDPKRENGRLPNPNSEPALCDPNLSFDAVTTVGN-KIFFKDRFFMLKVSERPTSV 318  
QY 84 ELISERKKNFPPSVDAFR-QGHNSVFLIKDKRWVYPPKKEKGYPKLIOD-EFPGIPS 141  
DB 319 NLISLSLWPTLPSSGIEAAYEIEARNOVFLFKDDKYWLISNLRPEPNYPKSIHSFGFNFVK 378  
QY 142 PLDAAYECHRGEQAGVLFQGHGHRNGTGHGNSHTHGPEVWRCSPHLVLSLTSNDHG 201  
DB 379 KIDAAV-----FNPRFYK----- 391QY 202 ATYAFSGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAAFSWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNDQYWRDERQMMDPGPKLITKNFGIGPK-IDAVFYKKNKYYFFQGSNOF 449  
QY 257 VF 258RESULT 10  
AAB91061 standard; Protein; 470 AA.

XX AAU91061;  
 AC 05-JUN-2002 (first entry)  
 DT Human head and neck tumour protein, SEQ ID NO 219.  
 DE Human; head and neck cancer; tumour; cytostatic; immunogenic; vaccine.  
 XX Homo sapiens.  
 OS WO200212329-A2.  
 XX 14-FEB-2002.  
 PD 01-AUG-2001; 2001WO-US24226.  
 PF 03-AUG-2000; 2000US-223281P.  
 XX 16-NOV-2000; 2000US-249933P.  
 XX (CORI-) CORIXA CORP.  
 PA Wang T, Fan L;  
 PI WPI; 2002-257467/30.  
 DR N-PSDB; ABRK54023.  
 XX Novel polynucleotide encoding head and neck tumour polypeptides, useful  
 PT in pharmaceutical compositions, e.g. vaccines, for treating head and  
 PS neck cancers -  
 XX Claim 2; Page 181-182; 200P; English.  
 CC The invention relates to an isolated polynucleotide (I) comprising  
 CC sequences selected from 273 sequences fully defined in the specification.  
 CC (I), including its encoded polypeptide (II), an antibody binding to (II),  
 CC a fusion protein comprising (II) and a T-cell population stimulated by  
 CC (I) or (II) are useful for stimulating an immune response in a patient  
 CC and treating head and neck cancer in a patient. An oligonucleotide (III)  
 CC that hybridises to (I) is useful for determining the presence of cancer  
 CC in a patient, by obtaining a biological sample from the patient,  
 CC contacting the sample with (III), detecting in the sample an amount of a  
 CC polynucleotide that hybridises to the oligonucleotide, and comparing the  
 CC amount of polynucleotide that hybridises to the oligonucleotide to a  
 CC predetermined cut-off value. (I) and (II) are useful in pharmaceutical  
 CC compositions, e.g. vaccines, and other compositions for the diagnosis and  
 CC treatment of head and neck cancer. AAU91061-AAU91062 represent human head  
 CC and neck cancer protein sequences of the invention.  
 CC  
 SQ Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 23; Length 470;  
 Best Local Similarity 27.3%; Pred. No. 3.9e-10;

Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNAVEGETPDDVTE--RCSGMSPDATLTDNGTMTLFPKGEFV-----KSHKMR 83  
 DB 260 SYGDPEKQRLPNPNSEPALCDPNLSPDAVTYGN-KIFFKDRFFWLKVSERKTSV 318  
 QY 84 ELISERKNKFPSPVDAAFR--QGHNSVFLIKGDKVWVYPEKKEKGYPKLQD-EFGGIS 141  
 DB 319 NLISSIMPTLPFGIEAAYEIBARNQVFLPKDDKVMYLSNLRBPNTPKSIHSGFPNFK 378  
 QY 142 PLDAVECHRGECQAEGLVFPQGHGRNGTGHGNSTHGPEYMRCSPHLVLSALTSDNHG 201  
 DB 379 KIDAAV-----FNPFRYR----- 391  
 QY 202 ATYAASGTHYWRDLTSRD---GHSWPIAHQWPGSPSAVDAAFSWEK-LYLVQSTQVY 256  
 DB 392 -TYFFVDNQYWRVDERRQMMDPGYPKLITKNFGIGPK-IDAVVFYSKNKYTFPQSSNPF 449  
 QY 257 VP 258

DB 450 EY 451

RESULT 11

ABP96800  
 ID ABP96800 standard; Protein; 470 AA.

AC ABP96800;

DT 05-JUN-2003 (first entry)

DE Human COPD related protein SEQ ID NO:50.

XX Human; chronic obstructive pulmonary disease; COPD; chronic lung disease.

OS Homo sapiens.

PN WO200297127-A2.

PD 05-DEC-2002.

PF 28-MAY-2002; 2002WO-EP05835.

PR 31-MAY-2001; 2001GB-0013266.

PA (FARB) BAYER AG.

PI Oellers N, Gehrmann M, Kallabis H, Hall R, Schulze T, Kroegel C;

DR WPI; 2003-140492/13.

XX N-PSDB; ACC46771.

PT Predicting, diagnosing or prognosing chronic lung disease, by detecting

PS a chronic obstructive pulmonary disease (COPD) gene in a biological

XX sample -  
 Claim 8; Page 153-155; 214pp; English.

CC The present invention describes a method for predicting, diagnosing or  
 CC prognosing chronic lung disease by detecting a chronic obstructive  
 CC pulmonary disease (COPD) gene related polynucleotide (see ACC46750 to  
 CC ACC46777, which encode the COPD related proteins in ABP96779 to  
 CC ABP96806). The method is useful for predicting, diagnosing or prognosing  
 CC chronic lung disease in a biological sample. The COPD genes and proteins  
 CC encoded by them from the present invention (I) can be used for treating  
 CC or preventing chronic lung disease in a mammal. (I) can be used in an  
 CC animal model for determining the efficacy, toxicity, or side effects of  
 CC treatment with (I), and determining the mechanism of action of (I).  
 CC ACC46778 to ACC46903 represent COPD related PCR primers and probes used  
 CC in an example from the present invention.  
 CC  
 SQ Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;  
 Best Local Similarity 27.3%; Pred. No. 3.9e-10;

Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNAVEGETPDDVTE--RCSGMSPDATLTDNGTMTLFPKGEFV-----KSHKMR 83  
 DB 260 SYGDPEKQRLPNPNSEPALCDPNLSPDAVTYGN-KIFFKDRFFWLKVSERKTSV 318  
 QY 84 ELISERKNKFPSPVDAAFR--QGHNSVFLIKGDKVWVYPEKKEKGYPKLQD-EFGGIS 141  
 DB 319 NLISSIMPTLPFGIEAAYEIBARNQVFLPKDDKVMYLSNLRBPNTPKSIHSGFPNFK 378  
 QY 142 PLDAVECHRGECQAEGLVFPQGHGRNGTGHGNSTHGPEYMRCSPHLVLSALTSDNHG 201  
 DB 379 KIDAAV-----FNPFRYR----- 391  
 QY 202 ATYAASGTHYWRDLTSRD---GHSWPIAHQWPGSPSAVDAAFSWEK-LYLVQSTQVY 256  
 DB 392 -TYFFVDNQYWRVDERRQMMDPGYPKLITKNFGIGPK-IDAVVFYSKNKYTFPQSSNPF 449

QY 257 VF 258  
DB 450 EX 451

## RESULT 12

ABR48203 standard; Protein; 470 AA.

ABR48203;

12-JUN-2003 (first entry)

Human bladder cancer associated protein sequence SEQ ID NO:125.

Human; bladder cancer; cytostatic; gene therapy; vaccine.

Homo sapiens.

WO2003003906-A2.

16-JAN-2003.

03-JUL-2002; 2002WO-US21338.

03-JUL-2001; 2001US-302814P.

03-AUG-2001; 2001US-310099P.

08-NOV-2001; 2001US-343705P.

13-NOV-2001; 2001US-350666P.

12-APR-2002; 2002US-372246P.

(EOSB-) EOS BIOTECHNOLOGY INC.

Mack DH, Aziz N;

WPI; 2003-201532/19.

N-PSDB; ACC51017.

Detecting a bladder cancer-associated transcript in a cell from a

patient, comprising contacting a biological sample from the patient with

a bladder cancer-associated polynucleotide or antibody

Claim 10, Page 275; 307pp; English.

The present invention describes a method for detecting a bladder cancer-

associated transcript in a cell from a patient. The method comprises

contacting a biological sample from the patient with a polynucleotide

that selectively hybridizes to a sequence that is 80 % identical to a

table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059

encode the human bladder cancer-associated sequences from the present invention

ABR48242). Bladder cancer-associated sequences from the present invention

have cytostatic activities, and can be used in antisense gene therapy and

in vaccine production. The method can be used for detecting a bladder

cancer-associated transcript in a cell from a patient. The method is

useful in diagnosing or treating bladder cancer and in screening for

compounds that modulate bladder cancer, such as hormones or antibodies.

The nucleic acid molecules from the present invention may be used in

various screening and diagnostic methods, and for gene therapy, vaccine

and/or antisense/inhibition applications.

Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10;

Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

30 SAHGNVAEGETKPDVTR--KSGDMSPDATLIDNGTMLFKGGEFVW---KSHKMDR 83

260 SLVGDPKRNQRLTNPNSBPALCDPNISFPAVTVGN-KIFPDKDFMVKSERPXTSV 318

84 ELISEMKNPSPSVDAFR-QGHNVSFLIKGDKRVYPPPKKKGKPKLQD-EFPGIPS 141

319 NLISIMPTLPSGIEAVERIARNQVFLFKDKXWLTLSINLRPPNPKSISFGFPMFVK 378

QY 142 PLDAVECHRGECQAGEVLFFQGHGRNGTGNSTHGPEYWRCSPHLVLSALTSNDHG 201  
DB 379 KIDAAV-----FNDRFYR----- 391

QY 202 ATYAFSGTHTYRLDTSRD---GMHSPIAHQMPQPSAVDAAFSBEK-LYVQGTQVY 256  
DB 392 -TYFVDNQYWRBYDERROMMDPGYPLITKNFGIGPK-IDAVFYKKNKYTFPQSNQF 449

QY 257 VF 258  
DB 450 EX 451

## RESULT 13

ABU56421 standard; Protein; 470 AA.

ABU56421;

02-APR-2003 (first entry)

Lung cancer-associated polypeptide #14.

Lung cancer-associated polypeptide; cytostatic; emphysema;

antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;

small cell lung cancer; benign lesion; precancerous lesion; bronchitis;

chronic obstructive pulmonary disease; hypersensitivity pneumonitis;

interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

Unidentified.

WO200286443-A2.

31-OCT-2002.

18-APR-2002; 2002WO-US12476.

18-APR-2001; 2001US-284770P.

10-MAY-2001; 2001US-290492P.

09-NOV-2001; 2001US-339245P.

13-NOV-2001; 2001US-350666P.

29-NOV-2001; 2001US-354370P.

12-APR-2002; 2002US-372246P.

(EOSB-) EOS BIOTECHNOLOGY INC.

Aziz N, Murray R;

WPI; 2003-093161/08.

N-PSDB; ABX76137.

Detecting a lung cancer-associated transcript in a cell from a patient

for treating lung cancer, by contacting a biological sample from the

patient with a polynucleotide that exhibits increased or decreased

expression in lung cancer

Claim 27, Page 198; 453pp; English.

The invention relates to a method for detecting a lung cancer-associated

transcript in a cell from a patient, comprising contacting a biological

sample from the patient with a polynucleotide that selectively hybridizes

to a sequence that is at least 80 % identical to a gene that exhibits

increased or decreased expression in lung cancer samples. Lung

cancer-associated polynucleotides and polypeptides are used for

identifying a compound that modulates a lung cancer-associated

polypeptide, for inhibiting proliferation of a lung cancer-associated

cell to treat lung cancer in a patient and for treating a mammal having

lung cancer by administering a modulatory compound identified. The

methods are useful for treating lung cancer, such as small cell lung

cancer, non-small cell lung cancer or other benign or precancerous

lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive

pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial

CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
 CC and polypeptides are useful for diagnostic purposes and as targets for  
 CC screening for therapeutic compounds that modulate lung cancer, such as  
 CC antibodies. Sequences ABUS6408-ABUS6745 represent lung cancer-associated  
 CC polypeptides of the invention.

XX Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
 Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAGEGTPDDVTE--RCSQSGSPDATLTDNGTMLPFKGFVW----KSHKMR 83  
 DB 260 SLYGDPKENVRLPNDNSPALCDPLSPDAVTYGN-KIFFPKRFFMLKXSERKTSV 318  
 QY 84 ELISERKNFSPVDAAFR-OCHNSVFLIKGDKWVYPEKKEGYKLLQD-EPPGIS 141  
 DB 319 NLISSLMPTLPSEIGEAAYEIRANQVFLPKDKYMLISLRPEPNPKSHSGFPNPK 378  
 QY 142 PLDAVECHRGECQAGVLPFGCHGRNGTGHGNSTHGPEYRCSPHLVLSALTSDNHG 201  
 DB 379 KIDAAV-----FNPRFYR----- 391  
 QY 202 ATYAASGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAAFSWEK-LVLYOQTOY 256  
 DB 392 -TYFPVDNQYMYRDERRQMDPGYKLLITKNFGIGPK-IDAVFYSKNKYTFPGSSNOP 449  
 QY 257 VF 258  
 DB 450 EY 451

RESULT 14  
 ABUS6663  
 ID ABUS6663 standard; Protein; 470 AA.

XX ABUS6663;

DT 02-APR-2003 (first entry)

XX Lung cancer-associated polypeptide #256.

XX Lung cancer-associated polypeptide; cytostatic; emphysema;  
 XX antiinflammatory; antiaesthetic; non-small cell lung cancer; atelectasis;  
 XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

XX Unidentified.

XX WO200286443-A2.

XX 31-OCT-2002.

XX 18-APR-2002; 2002WO-US12476.

XX 18-APR-2001; 2001US-284770P.

XX 10-MAY-2001; 2001US-290492P.

XX 09-NOV-2001; 2001US-339245P.

XX 13-NOV-2001; 2001US-350666P.

XX 29-NOV-2001; 2001US-334370P.

XX 12-APR-2002; 2002US-372246P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Aziz N, Murray R;

XX MPI; 2003-093161/08.

XX DR N-PSDB; ABX76392.

XX PT Detecting a lung cancer-associated transcript in a cell from a patient  
 for treating lung cancer, by contacting a biological sample from the

PT patient with a polynucleotide that exhibits increased or decreased  
 expression in lung cancer -

XX Claim 27; Page 385-386; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung  
 CC cancer-associated polynucleotides and polypeptides are used for  
 CC identifying a compound that modulates a lung cancer-associated  
 CC polypeptide, for inhibiting proliferation of a lung cancer-associated  
 CC cell to treat lung cancer in a patient and for treating a mammal having  
 CC lung cancer by administering a modulatory compound identified. The  
 CC methods are useful for treating lung cancer, such as small cell lung  
 CC cancer, non-small cell lung cancer or other benign or precancerous  
 CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive  
 CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial  
 CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
 CC and polypeptides are useful for diagnostic purposes and as targets for  
 CC screening for therapeutic compounds that modulate lung cancer, such as  
 CC antibodies. Sequences ABUS6408-ABUS6745 represent lung cancer-associated  
 CC polypeptides of the invention.

XX Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
 Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAGEGTPDDVTE--RCSQSGSPDATLTDNGTMLPFKGFVW----KSHKMR 83  
 DB 260 SLYGDPKENVRLPNDNSPALCDPLSPDAVTYGN-KIFFPKRFFMLKXSERKTSV 318  
 QY 84 ELISERKNFSPVDAAFR-OCHNSVFLIKGDKWVYPEKKEGYKLLQD-EPPGIS 141  
 DB 319 NLISSLMPTLPSEIGEAAYEIRANQVFLPKDKYMLISLRPEPNPKSHSGFPNPK 378  
 QY 142 PLDAVECHRGECQAGVLPFGCHGRNGTGHGNSTHGPEYRCSPHLVLSALTSDNHG 201  
 DB 379 KIDAAV-----FNPRFYR----- 391  
 QY 202 ATYAASGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAAFSWEK-LVLYOQTOY 256  
 DB 392 -TYFPVDNQYMYRDERRQMDPGYKLLITKNFGIGPK-IDAVFYSKNKYTFPGSSNOP 449  
 QY 257 VF 258  
 DB 450 EY 451

RESULT 15

XX AAB43772  
 ID AAB43772 standard; Protein; 473 AA.

XX AAB43772;

XX 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO.1217.

XX Human; cancer associated gene; cancer antigen; detection; cancer;  
 XX diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 XX antidiabetic; antiaesthetic; antineumatic; antiarthritic; antiviral;  
 XX antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;  
 XX dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;  
 XX vasotropic; antiporiatic; antiangiogenic; gene therapy; inflammation;  
 XX immune disorder; hematopoietic cell disorder; autoimmune disorder;  
 XX allergic reaction; graft versus host disease; organ rejection;  
 XX haemostatic; thrombolytic; cardiovascular disorder; infection;  
 XX neurological disease; drug screening.



OS Homo sapiens.  
XX  
PN W0200055350-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000MO-US05882.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
XX WPI; 2000-587533/55.  
DR N-PSDB; AAC77981.  
XX  
XX  
PT Novel isolated nucleic acids comprising sequences encoding peptides  
XX useful for treating or diagnosing e.g. cancer -  
XX  
XX  
PS Claim 11; Page 1848-1850; 2352pp; English.  
XX  
CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
CC antidiabetic; antihypertensive; antineoplastic; antiarthritic;  
CC antiinflammatory; antihypertensive; antidiabetic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;  
CC neoplastic; vasotropic; antiproliferative and antiangiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancer, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 473 AA;  
XX  
Query Match 9.3%; Score 202.5; DB 21; Length 473;  
Best Local Similarity 27.3%; Pred. No. 4e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
XX  
QY 30 SAHGNVABGRTKPDVTE--RCSDEMSPDATLIDNGTMLFFKGEFVW---KSHKMDR 83  
DB 263 SLYGDPKRNQRLPNPNSSEPLCDPLSLFPAVTVGN-KIFFKDRFFMLKXSERPKTSV 321  
QY 84 ELISERKWKFPSPVDAFR-QGNSVFLIKGDKVWYPPPEKKEGYPKLIQD-EPPGIPS 141  
DB 322 NLISSLMLPTLPSEIAEAYEIAENQVFLFPDCKYMLISMLRPPNYPKSIHSRGPFPYK 381  
QY 142 PLDAAYEGRGEGQAGVLFPGHGRNNGHNSTHHGPEYMRCSPHLVLSALTSNDNG 201  
DB 382 KIDAAV-----FNRFRYR----- 394  
QY 202 ATYAFSGTHYWLDSRD---GMSWPIAHQWPOGSAVDAFAFSWEK-LYLVQGTQVY 256  
DB 395 -TYFFVDNQYWRDERQMMDEPGLITKNGIGPK-IDANFYSKNTYYTFFQSGNQF 452  
QY 257 VF 258  
DB 453 EY 454

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 06:46:44 ; Search time 28 seconds  
(without alignments)  
590.841 Million cell updates/sec

Title: US-09-900-448-2

Perfect score: 2185

Sequence: 1 MARVLGAPVAGLWLSLCSL.....NAKALPQPNQNTSLIGCTH 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	202.5	9.3	470	3	US-08-068-392-2
2	202.5	9.3	470	4	US-08-396-988-2
3	202.5	9.3	470	4	US-09-391-104-26
4	199	9.1	517	4	US-09-391-104-32
5	189	9.1	519	3	US-09-211-704-7
6	189.5	8.7	476	3	US-08-704-711A-21
7	189.5	8.7	476	3	US-08-448-489-14
8	189.5	8.7	476	4	US-09-521-220-21
9	189.5	8.7	476	4	US-09-391-104-22
10	183.5	8.4	444	1	US-09-178-002-2
11	183.5	8.4	467	1	US-09-178-002-4
12	183.5	8.4	467	4	US-09-391-104-24
13	183.5	8.4	468	3	US-08-448-489-13
14	181.5	8.3	469	4	US-09-391-104-23
15	181.5	8.3	492	1	US-07-794-393-4
16	181.5	8.3	492	1	US-08-001-711-4
17	179.5	8.2	488	1	US-07-794-393-2
18	179.5	8.2	488	1	US-08-001-711-2
19	179.5	8.2	488	3	US-08-704-711A-22
20	179.5	8.2	488	4	US-09-521-220-22
21	179.5	8.2	488	4	US-09-391-104-31
22	179.5	8.2	489	3	US-08-448-489-11
23	177	8.1	477	3	US-08-704-711A-20
24	177	8.1	477	3	US-08-448-489-15
25	177	8.1	477	3	US-08-281-313-1
26	177	8.1	477	4	US-09-521-220-20
27	177	8.1	477	4	US-09-391-104-21

28	176.5	8.1	469	3	US-08-704-711A-16	Sequence 16, Appl
29	176.5	8.1	469	3	US-08-448-489-12	Sequence 12, Appl
30	176.5	8.1	469	4	US-09-521-220-16	Sequence 16, Appl
31	176	8.1	471	4	US-09-391-104-25	Sequence 25, Appl
32	175	8.0	466	3	US-08-704-711A-17	Sequence 17, Appl
33	175	8.0	466	4	US-09-521-220-17	Sequence 17, Appl
34	172	7.9	564	3	US-09-211-704A-8	Sequence 8, Appl
35	172	7.9	669	3	US-08-704-711A-3	Sequence 3, Appl
36	172	7.9	669	4	US-09-521-220-3	Sequence 3, Appl
37	172	7.9	669	4	US-09-391-104-29	Sequence 29, Appl
38	169.5	7.8	508	4	US-09-391-104-18	Sequence 18, Appl
39	167.5	7.7	411	4	US-09-171-545-3	Sequence 3, Appl
40	167.5	7.7	416	4	US-09-171-545-4	Sequence 4, Appl
41	167.5	7.7	508	4	US-09-171-545-1	Sequence 1, Appl
42	164.5	7.5	324	2	US-08-816-755-2	Sequence 2, Appl
43	164.5	7.5	324	3	US-09-090-673-2	Sequence 2, Appl
44	161.5	7.4	532	3	US-09-294-841-2	Sequence 2, Appl
45	154	7.0	419	4	US-09-171-545-5	Sequence 5, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-068-392-2
; Sequence 2, Application US/08068392
; Patent No. 6150152
; GENERAL INFORMATION:
; APPLICANT: Shapiro, Steven M.
; TITLE OF INVENTION: Human Macrophage Metalloproteinase
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SM
; STREET: 800 N. Lindbergh Blvd.
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,392
; FILING DATE: 19930528
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyer, Scott J.
; REGISTRATION NUMBER: 25275
; REFERENCE/DOCKET NUMBER: 07-24(12406)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 694-3117
; TELEFAX: (314) 694-5435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-068-392-2

Query Match 9.3%; Score 202.5; DB 3; Length 470;
Best Local Similarity 27.3%; Pred. No. 2.4e-12;
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGVAVAGETPDVDYR--RCSDGMSPDATTLDDNGTMLPFKKEPFVW---KSHKMR 83
DB 260 SLVGPKEKQRLPNPNDSPPALCDPNLSFDATVTGN-KIFFKQRFVLKVSERPKSV 318
QY 84 ELISERWKNPSPVDAFR-QGHNSVFLIKGDKVWVPEKKEKGVPLKLD-EPFGIPS 141
DB 319 NLISLWPTLPSGIEAAVIEARNQVFLFKDKKWLISLRREPRYPKSHSFGFPNPK 378
```

QY 142 PLDAVECHRGECQABGVLFQGHGRNGTGHGNSTHHGPEYWRCSPHLVLSALTSDNIG 201  
| : : : : :  
Db 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD-----GMHSMPIAHQWPOGSAVDAAFSWEK-LYLVQGTQYV 256  
| : : : : :  
Db 392 -TYFFVNDQYWRDYDERKQWMDPGYPLITKNQGIQPK-IDAVFYSKNKYYFFQGSNOF 449  
| : : : : :  
QY 257 VF 258  
| : : : : :  
Db 450 EY 451

RESULT 2  
US-08-396-988-2  
; Sequence 2, Application US/08396988  
; Patent No. 6204043  
; GENERAL INFORMATION:  
; APPLICANT: Shapito, Steven M.  
; TITLE OF INVENTION: Human Macrophage Metalloproteinase  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SM  
; STREET: 800 N. Lindbergh Blvd.  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/396,988  
; FILING DATE: 01-MAR-1995  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/068,392  
; FILING DATE: 28-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyer, Scott J.  
; REGISTRATION NUMBER: 25275  
; REFERENCE/DOCKET NUMBER: 07-24(12406)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3117  
; TELEFAX: (314)694-5435  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 470 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-396-988-2

Query Match 9.3%; Score 202.5; DB 3; Length 470;  
Best Local Similarity 27.3%; Pred. No. 2.4e-12;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGETKDPDPVTE--RCSDCGSPDATLIDNGTMLPFKGEFVW-----KSHKMDR 83  
| : : : : :  
Db 260 SLYGDKENQRLPDPNSPALCDPMLSPDAVTVGN-KIFFPKDRFFMLKYSERKTSV 318  
| : : : : :  
QY 84 ELISERWKNPSPVDAFR-QGHSNVFLIKGDRVWYPPKKEKGYPKLLQD-BFPGIPS 141  
| : : : : :  
Db 319 NIISSLMPTLPISGIEAAYEIEARNQVFLFKODKXMYLISMLRBPBNPKSIHSFGFPNFK 378  
| : : : : :  
QY 142 PLDAVECHRGECQABGVLFQGHGRNGTGHGNSTHHGPEYWRCSPHLVLSALTSDNIG 201  
| : : : : :  
Db 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD-----GMHSMPIAHQWPOGSAVDAAFSWEK-LYLVQGTQYV 256  
| : : : : :  
Db 450 EY 451

Db 392 -TYFFVNDQYWRDYDERKQWMDPGYPLITKNQGIQPK-IDAVFYSKNKYYFFQGSNOF 449  
| : : : : :  
QY 257 VF 258  
| : : : : :  
Db 450 EY 451

RESULT 3  
US-09-391-104-26  
; Sequence 26, Application US/09391104  
; Patent No. 6399371  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Palduto, Michael T.  
; APPLICANT: Magnuson, Scott R.  
; APPLICANT: Morgan, Douglas W.  
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEINASE GENE,  
; TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
; TITLE OF INVENTION: OF USING SAME  
; FILE REFERENCE: 6073.US.PI  
; CURRENT APPLICATION NUMBER: US/09/391,104  
; PRIOR FILING DATE: 1999-09-07  
; PRIOR APPLICATION NUMBER: US 08/814,394  
; PRIOR FILING DATE: 1997-03-11  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-391-104-26

Query Match 9.3%; Score 202.5; DB 4; Length 470;  
Best Local Similarity 27.3%; Pred. No. 2.4e-12;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGETKDPDPVTE--RCSDCGSPDATLIDNGTMLPFKGEFVW-----KSHKMDR 83  
| : : : : :  
Db 260 SLYGDKENQRLPDPNSPALCDPMLSPDAVTVGN-KIFFPKDRFFMLKYSERKTSV 318  
| : : : : :  
QY 84 ELISERWKNPSPVDAFR-QGHSNVFLIKGDRVWYPPKKEKGYPKLLQD-BFPGIPS 141  
| : : : : :  
Db 319 NIISSLMPTLPISGIEAAYEIEARNQVFLFKODKXMYLISMLRBPBNPKSIHSFGFPNFK 378  
| : : : : :  
QY 142 PLDAVECHRGECQABGVLFQGHGRNGTGHGNSTHHGPEYWRCSPHLVLSALTSDNIG 201  
| : : : : :  
Db 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD-----GMHSMPIAHQWPOGSAVDAAFSWEK-LYLVQGTQYV 256  
| : : : : :  
Db 392 -TYFFVNDQYWRDYDERKQWMDPGYPLITKNQGIQPK-IDAVFYSKNKYYFFQGSNOF 449  
| : : : : :  
QY 257 VF 258  
| : : : : :  
Db 450 EY 451

RESULT 4  
US-09-391-104-32  
; Sequence 32, Application US/09391104  
; Patent No. 6399371  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Palduto, Michael T.  
; APPLICANT: Magnuson, Scott R.  
; APPLICANT: Morgan, Douglas W.  
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEINASE GENE,  
; TITLE OF INVENTION: OF USING SAME  
; FILE REFERENCE: 6073.US.PI  
; CURRENT APPLICATION NUMBER: US/09/391,104  
; CURRENT FILING DATE: 1999-09-07

PRIOR APPLICATION NUMBER: US 08/814,394  
 PRIOR FILING DATE: 1997-03-11  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO 32  
 LENGTH: 517  
 TYPE: PR  
 ORGANISM: Homo sapiens  
 US-09-391-104-32

Query Match 9.1%; Score 199; DB 4; Length 517;  
 Best Local Similarity 25.0%; Pred. No. 6.3e-12;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----NSLCWSLAATPL-----PPTSAHGNAAGETKPPDPVTERCS 51  
 DB 188 VDDPLRYGIPYEDKVRWOLYGVRESVSPTAQPBEPPPLPEPPDNRSSAPPRKDVPHRCS 247  
 QY 52 DQMSPDATTLDDNGTLPFKGEPFWKSHKMDRELISER-----WKNPP---SPVDAAF 101  
 DB 248 T--HFDVA--QIRGEAFPFKGYFWRLTR--DRHLVSLQPAQMRFRWGLPLHDSVDAY 303  
 QY 102 -RQHSNVLKGDQVYVPEKKEKGYKLLQDEFPPIPSPLDAAVECHRGSCQAEVYL 160  
 DB 304 ERTSDHKIVFFKGDVWPKDNNVEGYRPSV--FSLPPGIDAA----- 348  
 QY 161 FQGHCHRGTHGNSHTHGPEYMRCSPHLVLSALTSNHNATYAFSGTHYRLDT--SRD 219  
 DB 349 FSWAHNR-----TYFFKQDLVYMYDDHTH 374  
 QY 220 GHSWPI-AHQWPGSVAADAAFSWE--KLVLVQGTQVYVFLTKGGYTLVSGYKLEK 277  
 DB 375 MDGPYAGSPLRWGVSTLDDAMRMSDGASYFRG--QETWKVLDELVAAPGPOSTARD 433  
 QY 278 -----VGTPHGIILDSVDAA 292  
 DB 434 WLVCBDSQADGSVAAGVDAA 453

RESULT 5  
 US-09-211-704A-7  
 Sequence 7, Application US/09211704A  
 Patent No. 6271014  
 GENERAL INFORMATION:  
 APPLICANT: de Saint-Vie, Blandine Marie  
 APPLICANT: Fossiez, Francois  
 APPLICANT: Caux, Christophe  
 APPLICANT: Lebeque, Serge J.B.  
 TITLE OF INVENTION: Mammalian Proteases; Related Reagents  
 TITLE OF INVENTION: and Methods  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DNA Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/211,704A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/005,263  
 FILING DATE: 09-JAN-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,090

REFERENCE/DOCKET NUMBER: SF0781K  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650)852-9196  
 TELEFAX: (650)496-1200  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 519 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-211-704A-7

Query Match 9.1%; Score 199; DB 3; Length 519;  
 Best Local Similarity 25.0%; Pred. No. 6.4e-12;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----NSLCWSLAATPL-----PPTSAHGNAAGETKPPDPVTERCS 51  
 DB 190 VDDPLRYGIPYEDKVRWOLYGVRESVSPTAQPBEPPPLPEPPDNRSSAPPRKDVPHRCS 249  
 QY 52 DQMSPDATTLDDNGTLPFKGEPFWKSHKMDRELISER-----WKNPP---SPVDAAF 101  
 DB 250 T--HFDVA--QIRGEAFPFKGYFWRLTR--DRHLVSLQPAQMRFRWGLPLHDSVDAY 305  
 QY 102 -RQHSNVLKGDQVYVPEKKEKGYKLLQDEFPPIPSPLDAAVECHRGSCQAEVYL 160  
 DB 306 ERTSDHKIVFFKGDVWPKDNNVEGYRPSV--FSLPPGIDAA----- 350  
 QY 161 FQGHCHRGTHGNSHTHGPEYMRCSPHLVLSALTSNHNATYAFSGTHYRLDT--SRD 219  
 DB 351 FSWAHNR-----TYFFKQDLVYMYDDHTH 376  
 QY 220 GHSWPI-AHQWPGSVAADAAFSWE--KLVLVQGTQVYVFLTKGGYTLVSGYKLEK 277  
 DB 377 MDGPYAGSPLRWGVSTLDDAMRMSDGASYFRG--QETWKVLDELVAAPGPOSTARD 435  
 QY 278 -----VGTPHGIILDSVDAA 292  
 DB 436 WLVCBDSQADGSVAAGVDAA 455

RESULT 6  
 US-08-704-711A-21  
 Sequence 21, Application US/08704711A  
 Patent No. 6114159  
 GENERAL INFORMATION:  
 APPLICANT: WILLY, Horst  
 APPLICANT: HINZMANN, Bernd  
 TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
 TITLE OF INVENTION: METALLOPROTEASES, THEIR PRODUCTION AND USE  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/704,711A  
 FILING DATE: 20-NOV-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/DE95/00357  
 FILING DATE: 17-MAR-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE 4438838.1

FILING DATE: 21-OCT-1994  
PRIOR APPLICATION DATA: DE 4409663.1  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-704-711A-21

Query Match 8.7%; Score 189.5; DB 3; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;

Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

QY 25 PLPTSAHGVABGETKPDVTERCSDDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKWR 83  
DB 272 PLVPTK---SVPSGSEMP-----AKCDPALSPDAISTL--RGEYLFPRKRYFWRGRSHMP 321  
QY 84 E-----LISERWKNFSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120  
DB 322 EPEFHLSAFWPSLPYLDAAYEVNSRDYFIKGNFMAIRGNEVOAGYPRGIHTLGF 381  
QY 121 P-----EKK-----EKGPKLQDEFPGIPSPDLAAVE 148  
DB 382 PTIRKIDAAVSDKEKKKTYFFADKYWRPDENSQSMGQFPRIADDPGVEPKVAVL- 440  
QY 149 CHRGECAQAGVLPF 162  
DB 441 -----QAFGFYF 448

RESULT 7  
US-08-448-489-14  
Sequence 14, Application US/08448489  
Patent No. 6184022  
GENERAL INFORMATION:  
APPLICANT: SEIKI, Mochiharu  
APPLICANT: SATO, Hiroshi  
APPLICANT: SHINAGAWA, Akira  
TITLE OF INVENTION: NOVEL METALLOPROTEINASE AND ENCODING DNA THEREFOR  
FILE REFERENCE: 55-290P  
CURRENT FILING DATE: US/08/448, 489  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Known Member of  
OTHER INFORMATION: Matrix Metalloproteinase Family  
US-08-448-489-14

Query Match 8.7%; Score 189.5; DB 3; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

QY 25 PLPTSAHGVABGETKPDVTERCSDDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKWR 83  
DB 272 PLVPTK---SVPSGSEMP-----AKCDPALSPDAISTL--RGEYLFPRKRYFWRGRSHMP 321  
QY 84 E-----LISERWKNFSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120

DB 322 EPEFHLSAFWPSLPYLDAAYEVNSRDYFIKGNFMAIRGNEVOAGYPRGIHTLGF 381  
QY 121 P-----EKK-----EKGPKLQDEFPGIPSPDLAAVE 148  
DB 382 PTIRKIDAAVSDKEKKKTYFFADKYWRPDENSQSMGQFPRIADDPGVEPKVAVL- 440  
QY 149 CHRGECAQAGVLPF 162  
DB 441 -----QAFGFYF 448

RESULT 8  
US-09-521-220-21  
Sequence 21, Application US/09521220  
Patent No. 639348  
GENERAL INFORMATION:  
APPLICANT: WILZ, Horst  
HINZMANN, Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
METALLOPROTEASES, THEIR PRODUCTION AND USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/521,220  
FILING DATE: 08-Mar-2000  
CLASSIFICATION: <Unknown>  
21-OCT-1994  
17-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/704,711  
FILING DATE: <Unknown>  
APPLICATION NUMBER: DE 4438838.1  
FILING DATE: 21-OCT-1994  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 21:  
US-09-521-220-21

Query Match 8.7%; Score 189.5; DB 4; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

QY 25 PLPTSAHGVABGETKPDVTERCSDDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKWR 83  
DB 272 PLVPTK---SVPSGSEMP-----AKCDPALSPDAISTL--RGEYLFPRKRYFWRGRSHMP 321  
QY 84 E-----LISERWKNFSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120

```

; ORGANISM: Homo sapiens
US-09-178-002-2
; Sequence 2, Application US/09178002
; Patent No. H001973
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Falduto, Michael T.
; APPLICANT: Magnuson, Scott R.
; APPLICANT: Morgan, Douglas W.
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEINASE GENE,
; TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS
; FILE REFERENCE: 6073.US.P1
; CURRENT APPLICATION NUMBER: US/09/391,104
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: US 08/814,394
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-391-104-22
; Sequence 22, Application US/09391104
; Patent No. 6393371
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Falduto, Michael T.
; APPLICANT: Magnuson, Scott R.
; APPLICANT: Morgan, Douglas W.
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEINASE GENE,
; TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS
; FILE REFERENCE: 6073.US.P1
; CURRENT APPLICATION NUMBER: US/09/391,104
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: US 08/814,394
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-391-104-22

```

## RESULT 9

```

US-09-178-002-2
; Sequence 2, Application US/09178002
; Patent No. H001973
; GENERAL INFORMATION:
; APPLICANT: Hu, Shou-Ih
; TITLE OF INVENTION: Human Neutrophil Collagenase Splice Variant
; FILE REFERENCE: CGC 2048
; CURRENT APPLICATION NUMBER: US/09/178,002
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 444
; TYPE: PRT

```

```

Query Match
Best Local Similarity 28.4%; Score 189.5; DB 4; Length 476;
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

```

```

QY 25 PLPTSAHGVABGETKPDVTERCSGDSGSPDA-TTLDNGTMLFFKGEFVWVKSHKMDR 83
DB 272 PLVPTK---SVPSGSEMP-----AKCDPALSPALISTL-RGEVLPFKDKYFRKSHMNP 321
QY 84 E---LISERKVNPPSPVDAFR-QGHSVFLIKGDKVWV-----YP 120
DB 322 EPEFHLISAFWPSLPSTLDAAYEVNSRDVTFIFKGNFMAIRGNEVOAGYPRGIHTLGF 381
QY 121 P-----EKK-----EKYPKLQDEFFPIPSPLDAVE 148
DB 382 PTIRKIDAAVSDREKKTTFYFADKYWRFDENSQSWEGFPRLIADDPGVEBKVDVAVL 440
QY 149 CHRGECCOAGSVLFF 162
DB 441 -----QATGFFFYF 448

```

## RESULT 10

```

US-09-178-002-2
; Sequence 2, Application US/09178002
; Patent No. H001973
; GENERAL INFORMATION:
; APPLICANT: Hu, Shou-Ih
; TITLE OF INVENTION: Human Neutrophil Collagenase Splice Variant
; FILE REFERENCE: CGC 2048
; CURRENT APPLICATION NUMBER: US/09/178,002
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 444
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
US-09-178-002-2

```

```

Query Match
Best Local Similarity 24.4%; Score 183.5; DB 1; Length 444;
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

```

```

QY 23 ATPPLPTSAHGVABGETKPDVTERCSGDSGSPDA-TTLDNGTMLFFKGEFVWVKSHKMDR 79
DB 242 SNDIQPT-----GPSTPKP-----CDPSLTFDAITTL-RGEVLPFKDKYFRKSHMNP 286
QY 80 --KMDRELISERKVNPPSPVDAFR-QGHSVFLIKGDKVWVPPKKEGYPKLQD-B 135
DB 287 LQREVNPFISLFWPSLPSTLDAAYEDPDRDLIFLFGKNGYMLSGYDILQGYPKDISNYG 346
QY 136 FPGIPSPDLAAVECHRGECOAEGVLPFGQHGHRNGTGHGNSTHRGPEYWRCSPLVLVSL 195
DB 347 FPGSVQALDAV-----FYR----- 361
QY 196 TSDNHGATYAFSGTHYWRDLTSR---DGMHSWPIAHQWPGSAVDAAFSWEKLYLVQG 252
DB 362 ----SKTYFVNDDQFWRYNQRFWEPGYPXISGAFPGISKVDVAVFOQEHFVHVSF 416
QY 253 TQVYVP 258
DB 417 PRYIAP 422

```

## RESULT 11

```

US-09-178-002-4
; Sequence 4, Application US/09178002
; Patent No. H001973
; GENERAL INFORMATION:
; APPLICANT: Hu, Shou-Ih
; TITLE OF INVENTION: Human Neutrophil Collagenase Splice Variant
; FILE REFERENCE: CGC 2048
; CURRENT APPLICATION NUMBER: US/09/178,002
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-178-002-4

```

```

Query Match
Best Local Similarity 24.4%; Score 183.5; DB 1; Length 467;
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

```

```

QY 23 ATPPLPTSAHGVABGETKPDVTERCSGDSGSPDA-TTLDNGTMLFFKGEFVWVKSHKMDR 79
DB 265 SNDIQPT-----GPSTPKP-----CDPSLTFDAITTL-RGEVLPFKDKYFRKSHMNP 309
QY 80 --KMDRELISERKVNPPSPVDAFR-QGHSVFLIKGDKVWVPPKKEGYPKLQD-B 135
DB 310 LQREVNPFISLFWPSLPSTLDAAYEDPDRDLIFLFGKNGYMLSGYDILQGYPKDISNYG 369
QY 136 FPGIPSPDLAAVECHRGECOAEGVLPFGQHGHRNGTGHGNSTHRGPEYWRCSPLVLVSL 195
DB 370 FPGSVQALDAV-----FYR----- 384
QY 196 TSDNHGATYAFSGTHYWRDLTSR---DGMHSWPIAHQWPGSAVDAAFSWEKLYLVQG 252
DB 385 ----SKTYFVNDDQFWRYNQRFWEPGYPXISGAFPGISKVDVAVFOQEHFVHVSF 439
QY 253 TQVYVP 258
DB 440 PRYIAP 445

```

## RESULT 12

```

US-09-391-104-24

```

Sequence 24, Application US/09391104  
Patent No. 6393371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Palduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
FILE REFERENCE: OF USING SAME  
CURRENT APPLICATION NUMBER: US/09/391,104  
PRIOR FILING DATE: 1999-09-07  
PRIOR APPLICATION NUMBER: US 08/814,394  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 24  
LENGTH: 467  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-391-104-24

Query Match 8.4%; Score 183.5; DB 4; Length 467;  
Best Local Similarity 24.4%; Pred. No. 2.2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

QY 23 ATPLPPTSAHGNVABEETKDPDPTVTERCSGWSFDA-TTLDNGTMLFFKGEFVWKS-- 79  
DB 265 SNPIQPT-----GSPSTKP-----CDPSLTFDATTTL--RGEILFFDORFWMRRHPQ 309  
QY 80 --KMDRELISERKNPSPVDAAFRO-GHNSVFLIKGDKVWVYPPEKKEGYPKLQD-E 135  
DB 310 LQVEKNFISLFWPSLPTGIAAYEDPDRDLIFLFGNQWALSGYDILQGYKDISNYG 369  
QY 136 FPGIPSPDLAAVECHRGCOABGVLFQGHGHRNGTGHSNTHGPEYKCSPLVLSAL 195  
DB 370 FPGSPVQALDAV-----FYR----- 384  
QY 196 TSDNHGATYAFSGTHWRDLTSR---DGMHSPRIAHQWPGSPAVDAAFSWEKLYVQ 252  
DB 385 -----SKTYFFVNDQFWRDYNQRFMEPGYKISGAFPGIESKVDVAFQOEHPFHVFSG 439  
QY 253 TOYVVF 258  
DB 440 PRYIAF 445

RESULT 13  
US-08-448-489-13

Sequence 13, Application US/08448489  
Patent No. 6184022  
GENERAL INFORMATION:  
APPLICANT: SEIKI, Motoharu  
APPLICANT: SATO, Hiroshi  
APPLICANT: SHINAGAWA, Aizra  
TITLE OF INVENTION: NOVEL METALLOPROTEINASE AND ENCODING DNA THEREFOR  
FILE REFERENCE: 55-290P  
CURRENT APPLICATION NUMBER: US/08/448,489  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 468  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: X = UNKNOWN  
OTHER INFORMATION: Description of Unknown Organism: Known Member of  
OTHER INFORMATION: Matrix Metalloproteinase Family  
US-08-448-489-13

Query Match 8.4%; Score 183.5; DB 3; Length 468;  
Best Local Similarity 24.4%; Pred. No. 2.2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

QY 23 ATPLPPTSAHGNVABEETKDPDPTVTERCSGWSFDA-TTLDNGTMLFFKGEFVWKS-- 79  
DB 265 SNPIQPT-----GSPSTKP-----CDPSLTFDATTTL--RGEILFFDORFWMRRHPQ 309  
QY 80 --KMDRELISERKNPSPVDAAFRO-GHNSVFLIKGDKVWVYPPEKKEGYPKLQD-E 135  
DB 310 LQVEKNFISLFWPSLPTGIAAYEDPDRDLIFLFGNQWALSGYDILQGYKDISNYG 369  
QY 136 FPGIPSPDLAAVECHRGCOABGVLFQGHGHRNGTGHSNTHGPEYKCSPLVLSAL 195  
DB 370 FPGSPVQALDAV-----FYR----- 384  
QY 196 TSDNHGATYAFSGTHWRDLTSR---DGMHSPRIAHQWPGSPAVDAAFSWEKLYVQ 252  
DB 385 -----SKTYFFVNDQFWRDYNQRFMEPGYKISGAFPGIESKVDVAFQOEHPFHVFSG 439  
QY 253 TOYVVF 258  
DB 440 PRYIAF 445

RESULT 14  
US-09-391-104-23

Sequence 23, Application US/09391104  
Patent No. 6393371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Palduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
FILE REFERENCE: OF USING SAME  
CURRENT APPLICATION NUMBER: US/09/391,104  
PRIOR FILING DATE: 1999-09-07  
PRIOR APPLICATION NUMBER: US 08/814,394  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 23  
LENGTH: 469  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-391-104-23

Query Match 8.3%; Score 181.5; DB 4; Length 469;  
Best Local Similarity 26.3%; Pred. No. 3.6e-10;  
Matches 59; Conservative 26; Mismatches 82; Indels 57; Gaps 8;

QY 44 PVTERCSGWSFDA-TTLDNGTMLFFKGEFVWKSMDREL-----ISERKNPSPVD 98  
DB 272 PGPACDCKLFFDATTI--RGVWFFDORFVMTNPPYVRLNFIISVFPQLPNGL 329  
QY 99 AAFR-QHNSVFLIKGDKVWVYPPEKKEGYPKLQDEPFGIPSPDLAAVECHRGCOAB 157  
DB 330 AAYEPADRDEVPFFPKNKYMAVOGQVILHGYKDISYSP-GPP----- 371  
QY 158 GVLFFQGHGHRNGTGHSNTHGPEYKCSPLVLSALSDNHGATYAFSGTHWRDL-- 215  
DB 372 -----RTVKA-IDALISEENTGTYFFVANKTMYDEX 403  
QY 216 -TSRDGHSMPRIAHQWPGSPAVDAAFSWEKLYVQTOYVVF 258  
DB 404 KSMDBGPKYKMLAHDPFGIGHGVDAVFMKDGFFFPFGHGRQKF 447

RESULT 15  
US-07-794-393-4



```
; Sequence 4, Application US/07794393
; Patent No. 5236844
; GENERAL INFORMATION:
; APPLICANT: CHAMBERON, PIERRE
; APPLICANT: BASSET, PAUL
; APPLICANT: BELLOQ, JEAN-PIERRE
; TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Ave. NW Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/794,393
; FILING DATE: 1991.11.21
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9025326.1
; FILING DATE: 21-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1383.0040000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-794-393-4

Query Match      8.3%; Score 181.5; DB 1; Length 492;
Best Local Similarity 25.2%; Pred. No. 3.8e-10;
Matches 69; Conservative 26; Mismatches 86; Indels 93; Gaps 13;

QY 24 TELPPTSAH-----GNVAGETKPDVTERCSFGSDGSPDATTLDNGTMLPFKGEFVWKS 78
DB 270 SPAPLTSSQAGDTNTEIALLEPETPPDVCE-----SFDVVS-TIRGELFFKAGFVWRL 323
QY 79 HKWDR-----LISRWKNPSPVDAAFRQGNVFLIKGDKVWVYPPEKKEGYPKLQ 133
DB 324 RSGRLQPGYPALASRMWGLPSFVDAAFDQAQIWFQQAQYWVYDGEKPVLGPAFLSK 383
QY 134 DEFPGLPSPLDAVECHRECCQAEGLVLFQGHGHRNGTGHGNSTHGPEYMRCSPLVLS 193
DB 384 LGLQG--SPVHAALV-----WGPR-----400
QY 194 ALTSNDHGATVAFSGTHVWRLD--TSR-DGWSWPIAHQ--WPGPSAVDAAFSWEKTL 247
DB 401 -----KNKIYFFRGADYRFRHRTQVND-----PVPRRSTDMRGVPSRIDAAE-----444
QY 248 YLVQGTQYVYFLTKG-----GYTLVSGYPK 272
DB 445 ---QDAEGYAYFLRCHLYWKFDVVKVLEGFPR 475
```

Search completed: December 16, 2003, 06:55:24  
Job time : 29 secs

THIS PAGE BLANK (USPTO)

1	937.5	42.9	254	4	Q9B319	Q9B319 homo sapien
2	689.5	31.6	276	11	Q8K1U6	Q8K1U6 mus musculus
3	625	28.6	427	13	Q8UJPD	Q8UJPD oryzae lat
4	464.5	21.3	439	13	Q90WF7	Q90WF7 cyprinus car
5	456	20.9	261	13	Q90WM3	Q90W33 gallus galli
6	447.5	20.5	445	13	Q90310	Q90310 carassius a
7	445.5	20.4	430	13	Q8UJPD	Q8UJPD oryzae lat
8	442	20.2	446	13	P798Z5	P798Z5 oncorhynch
9	271	12.4	211	13	Q9DFR1	Q9DFR1 gillichthys
10	224	10.3	416	13	Q9DFR1	Q9DFR1 oncorhynch
11	200.5	9.2	453	13	Q12945	Q12945 gallus galli
12	199	9.1	634	4	Q8IWC3	Q8IWC3 homo sapien
13	196.5	9.0	240	4	Q93745	Q93745 homo sapien
14	196	9.0	464	6	Q93TV5	Q93TV5 cryoclae
15	189.5	8.7	393	6	Q8M118	Q8M118 felis silve
16	183.5	8.4	359	13	Q90331	Q90331 columba co

17	181.5	8.3	232	4	Q96D24	Q96d44 homo sapien
18	180.5	8.3	458	11	P97568	P97568 rattus norv
19	177	8.1	462	13	Q91956	Q91956 salmo salar
20	175.5	8.0	452	6	Q91T82	Q91c22 canis famli
21	175.5	8.0	478	11	Q62905	Q62905 rattus norv
22	175	8.0	567	5	Q9GTC3	Q9GTC3 drosoephila
23	175	8.0	570	5	Q8MTN6	Q8MTN6 drosoephila
24	175	8.0	584	5	Q8MT35	Q8MT35 drosoephila
25	175	8.0	613	5	Q9W122	Q9W122 drosoephila
26	173.5	7.9	656	13	Q8UW23	Q8UW23 xenopus lae
27	173	7.9	479	11	Q922W6	Q922W6 mus musculus
28	173	7.9	657	13	Q9P7U7	Q9P7U7 oryzaiae lat
29	172.5	7.9	471	13	Q98859	Q98859 cynops pyrr
30	172.5	7.9	546	4	Q988F2	Q988f2 oryzaiae lat
31	172	7.9	546	4	Q98B96	Q98b96 homo sapien
32	169.5	7.8	478	4	Q9BSH7	Q9BSH7 homo sapien
33	169.5	7.8	521	5	O61266	O61266 caenorhabdi
34	166.5	7.6	478	11	Q91X32	Q91X32 mus musculus
35	165	7.6	378	13	Q93363	Q93363 gallus galli
36	164.5	7.5	472	13	Q93342	Q93342 gallus galli
37	164	7.5	464	11	Q9EBL5	Q9EBL5 mus musculus
38	164	7.5	655	13	Q9W635	Q9W635 oncorhynchus
39	162.5	7.4	478	11	Q8V114	Q8V114 mus musculus
40	162	7.4	492	13	Q8OCW9	Q8OCW9 oryzaiae lat
41	161.5	7.4	401	6	O77765	O77765 bos taurus
42	161.5	7.4	478	11	Q9D080	Q9D080 mus musculus
43	159.5	7.3	465	11	Q8C209	Q8C209 mus musculus
44	158	7.2	388	6	P79272	P79272 sus scrofa
45	158	7.2	431	13	Q98TF3	Q98TF3 oryzaiae lat

## ALIGNMENTS

## RESULT 1

ID	Q9BS19	PRELIMINARY;	PRT;	254	AA
----	--------	--------------	------	-----	----

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE Similar to hemopexin.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC005395; AAH05395.1; --  
DR HSSP; P20058; IHNX.  
DR InterPro; IPR000585; Hemopexin.  
DR Pfam; PF00045; hemopexin\_3.  
DR SMART; SM00120; HK; 3.  
DR PROSITE; PS00024; HEMOPEXIN; 1.  
DR PROSITE; 254 AA; 28630 MW; C935C273YF6B21BA CRC64;  
SQ SSQUCNCE

Query Match	42.9%	Score 937.5;	DB 4	Length 254;
Best Local Similarity	71.3%	Pred. No. 1.1e-70;		
Matches 176; Conservative	0;	Mismatches 0;	Indels 71;	Gaps 1.

QY	1	MARVIGAPALGNSICSLAIATLPPTSAHGNVABEETPPDVTYERCSDGSPATT	60
Db	1	MARVIGAPALGNSICSLAIATPLPTSAHGNVABEETPPDVTYERCSDGSPATT	60
QY	61	LDDNGTMLPFKGEFPMWSHKMDRELISRMKNPSPVDAAIROGHSNVPLIKGDKWVYP	120
Db	61	LDDNGTMLPFKGEFPMWSHKMDRELISRMKNPSPVDAAIROGHSNVPLIKGDKWVYP	120
QY	121	PEKKEGVYKLLIDEPFGIPSPDLAAVECHRGECQAEAGVLPFO-----	163

```

Db 121 PEKKEKGYPKLQDEFPGISPLDAVAECHRGEQAGVLFPGQDREMFMDLATGTWKR 180
Qy 164 -----GHHRN 169
Db 181 SWPVGNCSSALRWLGRYCYFQGNQFLRDPVGEVPPRRYPRDYVFMPCPGHGHNRN 240
Qy 170 GTGHGNS 176
    |||||
Db 241 GTGHGNS 247

RESULT 2
Q8K1U6 PRELIMINARY; PRT; 276 AA.
AC Q8K1U6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hemopexin (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Alam J., Smith A.;
RT "Nucleotide sequence of the mouse haemopexin gene."
RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; X56829; CAA40160.1; -.
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF00045; hemopexin; 3.
DR SMART; SM00120; HX; 3.
FT NON TER 276
SQ SEQUENCE 276 AA; 31047 MW; BD336C649B2B767 CRC64;

Query Match 31.6%; Score 689.5; DB 11; Length 276;
Best Local Similarity 53.0%; Pred. No. 7.8e-50;
Matches 149; Conservative 12; Mismatches 41; Indels 79; Gaps 6;

Qy 1 MARVLGAPVALGMSLCWSLAATPLPPTSAHGNVAEGE--TKRDPVTERCSDGMSFDA 58
Db 1 MARTAVLVNLTIVLGLCWSLAVASPLP--IAHGVAEVENGTCKDSDVPEHCLDTWGFDA 58
Qy 59 TLLDDNGTMLFFGGEFVWKSHKMDRELISRWKKNFSPVDAAPFQGNSTVFLKGDVWV 118
Db 59 ATMDHNGTMLFFGGEFVWKSHKMDRELISRWKKNFSPVDAAPFQGNSTVFLKGDVWV 117
Qy 119 YPEPEKEKGYPKLQDEFPGISPLDAVAECHRGEQAGVLFPGQDREMFMDLATGTWKR 165
Db 118 YPEPEKEKGYPKLQDEFPGISPLDAVAECHRGEQAGVLFPGQDREMFMDLATGTWKR 177
Qy 166 -----GH-- 167
Db 178 ERSWPAVGNCTAALRWLEERYCYFQGNKFLRFPNVTGEVPPRRYPRDYVFMPCPGHGHNRN 237
Qy 168 RNSGTGHGNSHTHGPRVRCSPHVLVSLSDNNGATYAPS 207
Db 238 PRNGTAHGNSTH--PMHSRCSPPDGLSLSDHNGATYAPT 276

RESULT 3
Q8J1P8 PRELIMINARY; PRT; 427 AA.
AC Q8J1P8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Warm-temperature-acclimation-related-65 kDa-protein-like-protein.
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;

```

```

OC Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=HNI; TISSUE=Liver;
RC Hirayama M., Kobiyama A., Kinoshita S., Watabe S.;
RT "Primary structural analysis of medaka Wap65 and Wap65-like protein
RT and their gene expression in association with temperature acclimation
RT and during ontogeny."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB075199; BAB97304.1; -.
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF00045; hemopexin; 3.
DR SMART; SM00120; HX; 7.
SQ SEQUENCE 427 AA; 48000 MW; DED6B6A5BND38EB CRC64;

Query Match 28.6%; Score 625; DB 13; Length 427;
Best Local Similarity 33.0%; Pred. No. 3.5e-44;
Matches 146; Conservative 66; Mismatches 136; Indels 94; Gaps 12;

Qy 15 SLCSLSLAATPLPPTSAHGNVAEGEIK-PDPVTERCSDGMSFDAITLDDNGTMLFFGGE 73
Db 11 ALMLALTTRAAPLEDSAA---GGDSALPDRGA-----GIEPDATTPDDKQGTFFPKD 60
Qy 74 FVWKSHPMDRELISRWKKNFSPVDAAPF-----QG-HNSVFLKGDVWVYPPEKKEK 126
Db 61 HWKKGREGDAQPSQYFKELNGHVDAAPRKNHPENQGDHDIYLFDDKVFSPFHTBLE 120
Qy 127 GYPKLLQDEFPGISPLDAVAECHRGEQAGVLFPGQDREMFMDLATGTWKR 164
Db 121 GYPKELQEDFPGPVPAHLDAVECPKGEVTVSVLPFKGPDVHYVDIVTKVTKTPHLP 180
Qy 165 -----HGR-----NGTGHGNSHTHGEY-MRC-----SP 188
Db 181 ACTSVRMLEHYCYCFGHNFTRFPQVYTGVTGYPDARRYEFCRCPDFGHGERTLKCS 240
Qy 189 HVLVSLTSDNNGATYAPSCTHYWRDLDSRDGWSHPIAHQWPGSAVDAAPFMEKTY 248
Db 241 DFRMDAITTDDTKRMTFKGSNYMRDLTRDGHAPITTSKELNGVDAYVSNDRIY 300
Qy 249 LVQGTGVVYFLTKGYTLVSGYPKRLEKEVTPHGIILDSVDAAPFCPSGSRHINAGR 308
Db 301 LKNGDQVYIYKAGAHPTLLEGPYPTLKEELN-----IEQVDAAPFCPGQRTAHIIQGRK 355
Qy 309 LMTLIDKSGAQAATWTLPLPHKEVQCALCMESKLSGNSANGPGLYLIHGPNLYCYSDV 368
Db 356 IYIYNLAATPRBITLDAAPLPFGDIDAAFC-----SSDGKIKFGSNYHYHDS 403
Qy 369 EKLNAAKALPOPOVNTS-LIGC 389
Db 404 MLVMSRIAPILAKTISAMVGC 425

RESULT 4
Q90WF7 PRELIMINARY; PRT; 439 AA.
AC Q90WF7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Warm-temperature-acclimation-related-65 kDa-protein precursor.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RA Kinoshita S., Itoi S., Watabe S.;
RT "cDNA Cloning and Characterization of the Warm-Temperature-
RT Acclimation-Related-65 kDa-Protein."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

```

DR	EMBL; AB052623; BAB60809.1; -.
DR	InterPro; IPR000585; Hemopexin.
DR	Pfam; PF00045; hemopexin; 5.
DR	SMART; SM00120; Hx; 5.
KM	Signal.
FT	SIGNAL
FT	CHAIN
SO	SEQUENCE
	1 22 POTENTIAL.
	23 439 POTENTIAL.
	439 AA; 50045 MW; 24CFB96629ABA9D CRC64;
Query Match	21.3%; Score 464.5; DB 13; Length 439;
Best Local Similarity	29.1%; Pred. No. 1.1e-30;
Matches 136; Conservative	69; Mismatches 139; Indels 123; Gaps 18;
Oy	SLCWSIATATL-----PPTSAAGNAVEGTRKDPVTRSCSGMSFDTLTLDNGTM 67
Dd	TLLCALILSPASSSVADDDPTAGHKPELHNEAKDL-----RCA-GMERDALAVEGIP 59
Oy	LFPKGEPYWKSHKNDRELISERKNP-----SPVDAAFR-----QGSNSVFILK 112
Dd	YFPKGDDHLFKGFPHGAELSN-----TFPELDDHNHGVDAAFRMSSEDSRDHDHQFFFL 116
Oy	GDKWAVYPERKEKKYRPKLQDERFGISPLDAVECHRGSCQAEGVLFPQG----- 164
Dd	DNKFVSYYKHLEKQYPKDIDSLFPGIEPDHDAAVECEPKDCTDDTVIFFKDEDIYHNNM 176
Oy	165 -----HGHR-----NQTGNSHTHOPEX-MRCS 187
Dd	KTQKVDEKFKSMPCNTGAFRYMEHYECFHGHQSKFPDVTGDVGKTPKETRTDFMRC- 235
Oy	188 PH-----LVLSALTSDNHGATYAFSGTHWRRLDTSRDMHSWPFIHQW 230
Dd	PHFGKSTEHIERQCGRVHLDATISDDDSIVAFRGHF--VSITDGKPHSDVESAFA 293
Oy	POGSAVDAASWEKLYLVOSTGYVFLTKGGTYLVSGYPRKLEKEVGTPHGIILDSVD 290
Dd	KELSEVDVAVSYEBEHLTMIKONEFYVKVGEPTHLETGPRLKEXVLG----TEGYPD 348
Oy	AAPICGSSRLIHMAGRLLMWLDLSGAQA-----TWTELPPHHKVDGALCMESLSPN 345
Dd	AAFCVADNHIAHVIGQTVYDELKATPRAPAKEGITTF---KKIDAMC-----GPK 399
Oy	346 SCSANGPGLYLINGENLYCYSDVEKLNAAKALPOPONT-SILAGCTH 391
Dd	400 -----GVTAVIGNHYLLDYDSPKIMMAKIMPEGRHVSQGLFGCDH 439
RESULT 5	
O90MR3	PRELIMINARY; PRT; 261 AA.
AC	O90MR3;
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
OS	Hemopexin (Fragment).
OC	Gallus gallus (Chicken).
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	TISSUE=Liver;
RA	Weisbach L.; Plant P.; Grieninger G.;
RT	"Chicken hemopexin gene expression and sequence of a partial length
RT	cDNA."
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; M37319; AL29887.1; -.
DR	InterPro; IPR000585; Hemopexin.
DR	Pfam; PF00045; hemopexin; 3.
DR	SMART; SM00120; Hx; 3.
PT	NON TER
SO	SEQUENCE
	261 AA; 29385 MW; 8C2BBE7010F3C9A0 CRC64;

Query Match	20.94	Score 456	DB 13	length 261
Best Local Similarity	36.24	Pred. No. 2, 98-30		
Matches 108	Conservative 46	Mismatches 92	Indels 52	Gaps 11
QY	96	PVDAAFRQGNHSVFLIKGDKVWYTPREKKE--KGYPKLLODFPGI99PLDAVECHERGE	153	
DB	10	PCBAALR-WLERYCYLGSTQFYRRFRPMSWEVLPCYPRDLRYPF--IRPP-----	55	
QY	154	QDAEGVLPFOGHGHRNGTGHGNSYTHHGPREYWCSPHVL5AL7SDNHCATYA6SGTHYMR	213	
DB	56	-----GRGHR-----HGN-TSMNAGDRCGE--PFAITSDDDSGH1AFAFGSLR	99	
QY	214	LDTSRDGMHSWPIAHQPPQGSAAVDAAFSMEKKLYLVQGTQVYFLT-KGGTYLVSGYRK	272	
DB	100	LDSWRDGMHAPQASHSPGLQGDVDAAF5MDKRYTLTQGSQVSYTVSGRGGHOLVEGYPK	159	
QY	273	RLEKEVSTPHGIILDSVDAAFICPSSRLHIMAGRIIMWLIDKSGAQTATWELPMPHEKV	332	
DB	160	ALDEGEVLP-----KADAAFTCGSAGELVYITDGMQRGVDTLSPRIADPQQLPYDGV	213	
QY	333	DGALCHEKSLGPNSSCANGPGLYIILHBPNLVCYSDVEKTLAAKALPQPNV-TSLGCG	389	
DB	214	DGAMC-----TADGTYLLRDRYHRHMDVALLAHPADPAPSLAVDLFHC	259	

	090310	PRELIMINARY;	PRT;	445 AA.
ID	Q90310			
AC	Q90310;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	MARM temperature acclimation-related 65-kDa protein precursor.			
OS	Carassius auratus (Goldfish).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;			
NCBI	Cyprinidae; Carassius.			
OX	NCBI_Taxid=7957;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skeletal muscle, and Hepatopancreas;			
RX	MEDLINE=95340486; Pubmed=7615502;			
RA	Kikuchi K., Yamashita M., Watabe S., Aida K.;			
RT	"The warm temperature acclimation-related 65-kDa protein, Wap65, in			
RL	goldfish and its gene expression.";			
RU	J. Biol. Chem. 270:17087-17092 (1995).			
RN	[2]			
RP	SEQUENCE OF 1-149 FROM N.A.			
RC	TISSUE=Hepatopancreas;			
RA	Kikuchi K., Watabe S., Aida K.;			
RL	Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; D50437; BA08928.1; -;			
DR	HSPB; D85428; BAA19835.1; -;			
DR	HSPB; P20058; IHXN.			
DR	InterPro; IPR00585; Hemopexin.			
DR	Pfam; PF00045; hemopexin; 5.			
DR	SMART; SM00120; HX; 4.			
FT	SIGNAL.			
FT	CHAIN	1 30	POTENTIAL.	
FT		31 445	WARM TEMPERATURE ACCLIMATION-RELATED 65-	
FT			KDA PROTEIN, WAP65.	
SEQ	SEQUENCE	445 AA; 50769 MW; 79C32EAB63BA82C6 CRC64;		
	Query March	20.5%; Score 447.5; DB 13; Length 445;		
	Best Local Similarity	28.6%; Pred.No. 2.9e-29;		
	Matches 132; Conservative	66; Mismatches 153; Indels 109; Gaps 16;		
DY	15 SLCSLAIAATPLPPTSAHGVAEGETPPDDPV-----TERCSGWSFATTLTDNGTML 68			
DY	:  :: :::			
DY	8 TLGLAASLSLAARPHNHKEDHYQODEPGQHGLHGHANLDRCG-GMEPDALVNBEGIPY 66			
DY	PFKGEFLWKSHKMDRELISERWKNFP-----SPVDAAFRQ-----GANSVFLLKG 113			
DY	:: :: ::			

```

Db 67 FPKDHLFGKHDAELSNB---TPRELDEHHHLGHVDAAFRMSSEDSRAHHDHQFFFLD 123
Qy 114 DKXWVYPERKEXKGYRLDDEPGISPLDAAVECHRGEOAGVLPFG-- 164
Db 124 TKVFSYKHLKEXDYPDISELFPGLPDHLDAAVECPDPCADTIIFFEGDEIYHDMK 183
Qy 165 -----HGR-----NGTGHGNSHTHGPY-MRCS 188
Db 184 TKVDEKPKSMRSTGAFRMDHYCPHGHQFKEPPIGEOGKPKXTRODTPMC-P 242
Qy 189 H-----LVLSALTSNHNATYAFSGTHYRLDTSRDGMSHPIAHQWP 221
Db 243 HFGQKTDEHIEBQCSRVLDAITSDDSVAFRGNH--LSITGDKFHSPTIESAFK 300
Qy 232 QGSAVDAAFSMEKLYLNGTOYVFLTKGYTLVSGYPRLEKEVGTGHIILDSVDA 291
Db 301 ELHSEVDAVFSYEGHLYMIDKNEVYVKGVPHTHLEGPRLKEVLG-----LEGVDA 355
Qy 292 AFICPGSSRLHIMAGRLMILDKSGAQTWELPMPH-EKVDGALCMKESLGPNSCSAN 350
Db 356 AFVCAADHNIHAYVGVYVDVLDKATPRVYKESINHLKIDVAMC-----GPK----- 405
Qy 351 GPGLYLHGNPLYCYSDVEKLNAAKALPQONT-SLIGCTH 391
Db 406 --GVTAVIGNHYYQFGSPMIMMAKIMPEQHRVSGLFGCDH 445

RESULT 7
ID Q8JIP9 PRELIMINARY; PRT; 430 AA.
AC Q8JIP9;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Warm-temperature-acclimation-related-65 KDa-protein.
GN MAP65.
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.
OC NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HNI;
RA Hirayama M., Kobiyama A., Kinoshita S., Watabe S.;
RT "Primary structural analysis of medaka Map65-like protein
RT and their gene expression in association with temperature acclimation
RT and during ontogeny."
RL Submitted (JBC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB075198; BAB97303.1;
DR InterPro: IPR000585; Hemopexin.
DR Pfam: PF00045; hemopexin; 2.
DR SMART: SM00120; HX; 4.
SQ SEQUENCE 430 AA; 49847 MW; A8BFC56B33AAEF15 CRC64;

Query Match 20.4%; Score 445.5; DB 13; Length 430;
Best Local Similarity 28.1%; Pred. No. 4e-29;
Matches 126; Conservative 67; Mismatches 155; Indels 101; Gaps 14;

Qy 16 LCWSLAIATPLPPTSAHGVNAEGETKPPDVTTERCSDGWSPDATLTDNGTMLFFKGEFV 75
Db 10 LCLALVLA---ADHHEHRRKGA-----VRDCK-GIEMDAVAVNEBGIPIYFKDHL 58
Qy 76 WKSHKMDRELISERKNRP-----SPVDAFR-----QGHNSVFLIKGDKVWVYPER 123
Db 59 FKGNHQAELSNKSFALDDHHLGHVDAAFRMYEDDLNHDMPFLDNKFAVYQHK 118
Qy 124 KEKGYRLDDEPGISPLDAAVECHRGEOAGVLPFG-- 164
Db 119 LEAGYPAISEVFPGLPDHLDAAVECPDPCADTIIFFEGDEIYHDMK 178
Qy 165 -----HGR-----NGTGHGNSHTHGPY-MRCS----- 187

```

```

Db 179 SMPNCTSAFRMEHYCFGHKSKFDPKTEVGRKYPKDAKFFMRCSKFEEDNDHEER 238
Qy 188 ---PHLVLSALTSNHNATYAFSGTHYRLDTSRDGMSHPIAHQWPQBSAVDAAFSWE 244
Db 239 ERSRVLDAITSDDSAGNIYAFRGNHYIRKDEBNDTLKDTIESAKELHSEVDAVFSYN 298
Qy 245 EKLTVQGVYVFLTKGYTLVSGYPRLEKEVGTGHIILDSVDAAFICPGSSRLHIM 304
Db 299 SHLYMKDQDLFYRGEPRHTLAGYPRKYVALG-----IKPILAAVYCDRIAHII 353
Qy 305 AGRRLMILDKSGAQTWELPMPH-HEKVDGALCMKESLGPNSCSANGELYLHGNLY 363
Db 354 KDRHMTVDVMSAFPRATNRPRISILKVDGAMC-----GPK-----GVKVRGNHY 401
Qy 364 CYSDEKLNAAKALPQONT-SLIGCTH 391
Db 402 HFSPKTFVAARALPQHRISLELFGCDH 430

RESULT 8
ID P79825 PRELIMINARY; PRT; 446 AA.
AC P79825;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hemopexin-like protein (fragment). (Salmo gairdneri).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OC NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97142135; PubMed=8988369;
RA Mot S., Duval J., Le Goff P.;
RT "Molecular cloning of a hemopexin-like cDNA from rainbow trout
RT liver."
RL DNA Seq. 6:311-318(1996).
DR EMBL: Z68112; CAA92147.1;
DR HSSP: P20058; IHNX.
DR InterPro: IPR000585; Hemopexin.
DR Pfam: PF00045; hemopexin; 4.
DR SMART: SM00120; HX; 4.
DR NON TER 1
SQ SEQUENCE 446 AA; 50454 MW; 13A8B4EA42B3727 CRC64;

Query Match 20.2%; Score 442; DB 13; Length 446;
Best Local Similarity 27.8%; Pred. No. 8.3e-29;
Matches 126; Conservative 64; Mismatches 168; Indels 96; Gaps 13;

Qy 16 LCWSLAIATPLPPTSAHGVNAEGETKPPD-----VTERCSDGWSPDATLTDNGTMLFF 70
Db 11 LCLVLAISHAHNHAHQGGDEBEGHGHGHLDRGCGIEMDAVAVTEBGIPIYF 69
Qy 71 KGFVWKSMDRELISERKNRP-----SPVDAFR-----GHNSVFLIKGDKVWVYPER 118
Db 70 KGHVFRKGFAGKALSNESFALDDHHLGHVDAAFRMYEDDLNHDMPFLDNKFAVYQHK 129
Qy 119 YPERKEXKGYRLDDEPGISPLDAAVECHRGEOAGVLPFG-- 164
Db 130 YKHQLETFPPKQISEVFPGLPDHLDAAVECPDPCADTIIFFEGDEIYHDMK 189
Qy 165 -----HGR-----NGTGHGNSHTHGPY-MRCS----- 187
Db 190 EKKFEGMNPNTSAFRMEHYCFGHQFSEKFDKTEBVGIRYKABDYPMKSGKGDIT 249
Qy 188 -----PHLVLSALTSNHNATYAFSGTHYRLDTSRDGMSHPIAHQWPQBSAVDA 239
Db 250 DHIERERCSRVLDAITSDAGNIYAFRGNHYLEODAGNDTMAADTIESDKELHSEVDA 309

```

QY 240 AFSGMEKLYVGGTGVYFLTKGGYTLVSGYPRKLEKEVGTGPIILDSVDAAFICPGSS 299  
 DB 310 TFSYENHLVWVKDKYIVYVGSHTLDSPKPLKEVLG-----VEGPIDAAFCVQDHI 364  
 QY 300 RLHIMAGRLIMLIDLSGGAQATWTELPW-HEKYDGLCMESKSLGPNCSANGPGLYLH 358  
 DB 365 IAHVHKQCYVDVLDLKSPVPVVEGSPFLFNKVDAMC-----GPE-----GVGLFK 412  
 QY 359 GPNLYCSDVEKLNAAKALPOP-ONVTSLLGCTH 391  
 DB 413 GNHYFHFQSVKVMMAKALPEEHKTLLEFGCDH 446

## RESULT 9

Q9DFN1 PRELIMINARY; PRT; 211 AA.  
 AC Q9DFN1; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hemopexin-like protein (Fragment).  
 OS Gallitichthys mirabilis (Long-jawed mudsucker).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Gobioidae;  
 OC Gobiidae; Gallitichthys.  
 NC NCB1\_TaxID=8222;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=21117151; PubMed=11172064;  
 RA Gracey A.Y., Troll J.V., Somero G.N.;

RT "Hypoxia-induced gene expression profiling in the euryoxic fish  
 RT Gallitichthys mirabilis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).  
 DR EMBL; AF266204; AAC13324.1; -.  
 DR HSSP; P20058; IHN.  
 DR InterPro; IPR000585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR SMART; SM00120; HX; 2.  
 FT NON TER 1  
 SQ SEQUENCE 211 AA; 24104 MW; EC149A2AC07B13CC CRC64;

Query Match 12.4%; Score 271; DB 13; Length 211;  
 Best Local Similarity 33.0%; Pred. No. 7.2e-15;  
 Matches 69; Conservative 37; Mismatches 83; Indels 20; Gaps 6;

QY 185 RCGPMLVLSALTSPDNHGAATAFSGTHYMRLDTSRDGHSMPDAHQPQSAVDAAFSWE 244  
 DB 21 KCS-RVHLDDITSDNAGNMTAFRPHNTEKDEGNDITVDNIEHAFELHSDVDAYSYD 79  
 QY 245 EKLTVGGTGVYFLTKGGYTLVSGYPRKLEKEVGTGPIILDSVDAAFICPGSSRLHIM 304  
 DB 80 DHLYMIDKHVHYKVAEPHTLHGGYKTYEELG-----IDGVDAAPVCEHDHINFI 134  
 QY 305 AGRLIMLIDLSGGAQATWTELPW-HEKYDGLCMESKSLGPNCSANGPGLYLHGNLY 363  
 DB 135 QGKNIFFVMDKVSREPRTNERPLALFKKVDAMC-----GPG-----GLKVVGNHFI 182  
 QY 364 CYSDVEKLNAAKALPOPONV-TSLLGCTH 391  
 DB 183 RFDQVLMVAGRALPEQHRVSTELFGCDH 211

## RESULT 10

Q9DFN1 PRELIMINARY; PRT; 116 AA.  
 AC Q9DFN1; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hemopexin-like protein variant 1 (Fragment).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Proteacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 NC NCB1\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bayne C.J., Gerwick L., Fujiki K., Nakao M., Yano T.;

RT "Immune-relevant (including acute phase) genes identified in the  
 RT livers of rainbow trout, Oncorhynchus mykiss, by means of suppression  
 RT subtractive hybridization.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF281339; AAG30014.1; -.  
 DR HSSP; P45452; IPEX.  
 DR InterPro; IPR000585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 1.  
 DR SMART; SM00120; HX; 1.  
 FT NON TER 1  
 SQ SEQUENCE 116 AA; 13379 MW; 4543D625B162B3BC CRC64;

Query Match 10.3%; Score 224; DB 13; Length 116;  
 Best Local Similarity 40.7%; Pred. No. 3e-11;  
 Matches 46; Conservative 20; Mismatches 35; Indels 12; Gaps 2;

QY 64 NGTMLPFKGEFVWKSHKMDRELISERKNPPS-----PVDAPR-----QGNHGVFLI 111  
 DB 1 NGHTFFKGDHMLNGCKGPAQVSSAFKELDYNHGHVDAAPRMNKEKPEEGHDITFR 60  
 QY 112 KSDKVVVPEPEKKEKGYPKLDDPPGIPSPDAAVECHGECAEGVLPFG 164  
 DB 61 LDDKVSYYNHSLEBGYPDIDLPFGVPSHVDAAVECPGECNSDVSVPFKG 113

## RESULT 11

Q12945 PRELIMINARY; PRT; 453 AA.

ID Q12945;  
 AC Q12945;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Vitroneurin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCB1\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=White Leghorn; TISSUE=Retina;  
 RX MEDLINE=98090066; PubMed=9362471;  
 RA Martinez-Morales J.R., Barbas J.A., Marti E., Bovolenta P., Edgar D.,  
 RA Rodriguez-Tebar A.;

RT "Vitroneurin is expressed in the ventral region of the neural tube and  
 RT promotes the differentiation of motor neurons.";  
 RL Development 124:5139-5147(1997).  
 DR EMBL; Y11030; CA471914.1; -.

DR HSSP; P45452; IPEX.  
 DR InterPro; IPR001212; Somatomedin\_B.  
 DR InterPro; IPR001212; Somatomedin\_B.  
 DR Pfam; PF00045; hemopexin; 4.  
 DR Pfam; PF01033; Somatomedin B; 1.  
 DR PRINTS; PRO00022; SOMATOMEDINB.  
 DR SMART; SM00120; HX; 4.  
 DR SMART; SM00201; SO; 1.  
 DR PROSITE; PS00024; HEMOPEXIN; 2.  
 DR PROSITE; PS00524; SOMATOMEDIN B; 1.  
 SQ SEQUENCE 453 AA; 51636 MW; 82B3BDDC2B9397C4 CRC64;

Query Match 9.2%; Score 200.5; DB 13; Length 453;  
 Best Local Similarity 25.8%; Pred. No. 1.6e-08;  
 Matches 91; Conservative 42; Mismatches 103; Indels 117; Gaps 20;

QY 26 LPPTSAHGNAVGEETRP---DPDVTGRCSDGMSFADATLDDNCTMLPFKGEFVWKSHKMD 82



```

Db      127 LHPITTTTTSDTRNPSLDDEPELCSRK-PFNAFTDLNKGSIYAFRGYFYLKSS 185
Qy      83 -----RELISERKWNKPSVDAAFR-----OGHNSVFLIKGDKVWVPEKKEKGYPKLQ 133
Db      186 VRPGYPLKISDW-GIGSPIDALFTRINCOG-KTYLFKSGQYRFPDGDALDPYIPDIS 242
Qy      134 DEFGIPLDAVECHRGECQAGVLFPQGHGRNGTGHNSTHGSPYWRCSPHVLVS 193
Db      243 EGFGIIPNDIDA-----PALPAH-----SYHGNE----- 267
Qy      194 ALTSDNHGATYAFEGTTHWRLDTSRDGHSWPIAHQMPQ-----GSAV-DAAF----- 241
Db      268 -----RYVFFKGYTW-----SYDFAHQPTQAECEKSSPSYFNHAFMRD 309
Qy      242 SWEKLYVVOGTQYVFLTKGQYTLVSGYPKLEKE--VGTPHGIILSDVDAFICPSS 299
Db      310 SWED-----IFLSFGSRMVGASSQRLISRWQGVNQ-IDAMAGRIYSSR 356
Qy      300 RLHIMAGR-----LM-WLDLKSQAQTWTLPMPEHEKVDGALC 337
Db      357 QPRRRSSRRHRKRYRNHRTNLGLMSWLN--SDSESTDSDW-----LSGSQC 403

```

## RESULT 12

```

Q81WC3 ID Q81WC3 PRELIMINARY; PRT; 634 AA.
AC Q81WC3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strauberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC040507; AA040507.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 634 AA; 70086 MW; 0D4A78663BAACAF CRC64;

```

Query Match 9.1%; Score 199; DB 4; Length 634;

Best Local Similarity 25.0%; Pred. No. 3.2e-08; Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

```

Qy      5 LGAPVALGL-----WSLCWSLAITPL-----PTSAHGNVAEGETKPPDPTVTERCS 51
Db      305 VGDPRLRGLEPYEDKVRWQLYGVRESVPTAQPEFPLBPBPNNRSAPPRKDVPHRCS 364
Qy      52 DGSFPAATLTDNGTMLPFKGEFVWKSMDRELISER-----WQNF--SPVDAF 101
Db      365 T--HFDVA-QIRGEAFFKGYFWRLTR--DRHLVSLQPAQMRFWRGFLPLHLSVDVAVY 420
Qy      102 -RGHNSVFLIKGDKVWVPPPEKKEKGYPKLQDEFGIISPLDAVECHRGECQAGVL 160
Db      421 ERSNDHKTIVFKGDRVWVFQNNVEBGYPKVPD-FSLPGGIDA----- 465
Qy      161 FFOGHGRNGTGHNSTHGSPYWRCSPHVLSLTSDNHGATYAFSGTHWRLDT--SRD 219
Db      466 FSAHADR-----TYFFDQLYWKYDHTRH 491
Qy      220 GHSWPI-AHQWPGSAVDAAPSWEK-KLYLVQGTQVYVFLTKGQYTLVSGYPKLEKE 277
Db      492 MDGYPKQSPKMRGVPSTLDDAMKMSDGAASYFRG-QEYWKVLDELVAFGYQSTARD 550
Qy      278 -----VGTPHGIILSDVDA 292
Db      551 WLVCGBSDQADGSAAGVDAA 570

```

## RESULT 13

```

Q99745 ID Q99745 PRELIMINARY; PRT; 240 AA.
AC Q99745;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Metalloproteinase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043200; PubMed=8226919;
RA Shapiro S.D., Kobayashi D.K., Ley T.J.;
RT "Cloning and characterization of a unique elastolytic
RT metalloproteinase produced by human alveolar macrophages."
RL J. Biol. Chem. 268:23824-23829(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97187267; PubMed=9034720;
RA Borden P., Heller R.A.;
RT "Transcriptional control of matrix metalloproteinases and the tissue
RT inhibitors of matrix metalloproteinases."
RL Crit. Rev. Eukaryot. Gene Expr. 7:159-178(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Lin D., Duncan M., Allen E., Araujo R., Aparicio A., Chai A.,
RA Chung E., Davis K., Federspiel N., Hyman S., Kaiman S., Kemp C.,
RA Kurd O., Lashari D., Lew H., Namath A., Oefner P., Roberts D.,
RA Heller R., Davis R.W.;
RT "Three Matrix Metalloproteinases on 81kb of Pl insert."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U78045; AAB36943.1; -.
DR HSSP; P39900; 1J1Z.
DR Interpro; IPR000585; Hemopexin.
DR Pfam; PF00045; hemopexin; 4.
DR SMART; SM00120; HX; 4.
DR PROSITE; PS00024; HEMOPEXIN; 1.
FT NON TER 1
SQ SEQUENCE 240 AA; 28406 MW; 3E620FAF17EB7CCC CRC64;

```

Query Match 9.0%; Score 196.5; DB 4; Length 240;

Best Local Similarity 26.9%; Pred. No. 1.5e-08; Matches 65; Conservative 22; Mismatches 92; Indels 63; Gaps 10;

```

Qy      30 SAHGNVAEGETKPPD--VTERCSDSGSPDATTLDNGTMLPFKGEFVW-----KSHKMDR 83
Db      30 SLVGDPKENGRLPNPNKSPALCDPMLSDAVTVGN-KIFFKXDFPMLKVSERPKTSV 88
Qy      84 ELISERKWNKPSVDAAFR-OGHNSVFLIKGDKVWVPPPEKKEKGYPKLQD-EFGIIS 141
Db      89 NLISLSMPTLPSSGIEAAYEIKANQVFLFKDDCYWYLSMLRPPNPKSISHSFGFPNFKV 148
Qy      142 PLDAVECHRGECQAGVLFPQGHGRNGTGHNSTHGSPYWRCSPHVLSLTSDNHG 201
Db      149 KIDAAV-----FNRFR----- 161
Qy      202 ATYAFSGTHWRLDTSRD-----GHSWPIAHQWPGSAVDAAPSWEK-LYLVQGTQVY 256
Db      162 -TYFFVQNYWRDERQWMDPEYPLIRKNGGIGPK-IDAVFYKKNKYYVFGQSNQF 219
Qy      257 VF 258
Db      220 EY 221

```

## RESULT 14

```

Q9TV75 ID Q9TV75 PRELIMINARY; PRT; 464 AA.

```

```

AC Q9TV75;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Matrix metalloproteinase-12 precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Japan White; TISSUE=long bones;
RX MEDLINE=98424349; PubMed=9753427;
RA Kobori M., Ikeda Y., Nara H., Kato M., Kumegawa M., Nojima H.,
RA Kawashima H.;
RT "Large scale isolation of osteoclast-specific genes by an improved
RT method involving the preparation of a subtracted cDNA library.";
RL Gene Cells 3:459-475(1998).
DR HBB; AB006779; BAA75922.1; -.
DR HBB; P39900.1; J17.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR006026; Nzn_MTPeptide.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase_M10_N; 1.
DR Pfam; PF03933; Peptidase_M10_N; 1.
DR PRINTS; PR00138; MATRIXIN.
DR SMART; SM00120; HK; 4.
DR SMART; SM00235; ZMNC; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR HydroLase; Metal-binding; Metalloprotease; Protease; Signal; Zinc.
KW SIGNAL
FT SIGNAL
SQ SEQUENCE 464 AA; 53585 MW; B9A81A391B24317E CRC64;

Query Match 9.0%; Score 196; DB 6; Length 464;
Best Local Similarity 27.4%; Pred. No. 3.8e-08;
Matches 64; Conservative 19; Mismatches 91; Indels 60; Gaps 8;

QY 30 SAHGNVAEGETKDPDVT--RCSGMSFPDATTLDNGTMLFKGEGFV-----KSHKMR 83
DB 255 SLVGEGEQPMKPNPEPTACDHLKDAVTVGN-KLFFKDSFFWKKPKSSTTSV 313
QY 84 ELISERWKNPSPVDAAFRQ--HNSVFLIKGRVWVYPPPEKKEGYPKLLQD--EPPGIPS 141
DB 314 RLISLIMPTLPSGIRAAVEIGDRHQVFLFKGDKFMLISHLRLQPNYPKSIHSLGFPDPVK 373
QY 142 PLDAAVECHRGECQAGVLFPGHGHNGTGHGNSTHHGPEYMRCSPHLVLSALTSDNHG 201
DB 374 KIDAAV-----FNPSL-----R 385
QY 202 ATYAAGSTHYWRDLTSR---DGNHWPFIHQWPGSPSAVDAAFSWEKLYLVQGT 252
DB 386 KTYFVDNLWKYRDERREVMADGYPKLLTHGPGIGPKIDAVYFYFQRYTFFQGS 439

```

RESULT 15

```

ID 08MT18 PRELIMINARY; PRT; 393 AA.
AC 08MT18;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Stromelysin-1 (Fragment).
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.;
RC TISSUE=Fibrosarcoma;

```

```

RA Sorensen K.C., Kitchell B.E.;
RT "Comparison of feline Stromelysin-1 (MMP3) cDNA with other known MMP3
RT cDNA sequences.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033627; AAK64605.1; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR006026; Nzn_MTPeptide.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF00045; hemopexin; 3.
DR Pfam; PF00413; Peptidase_M10_N; 1.
DR Pfam; PF03933; Peptidase_M10_N; 1.
DR PRINTS; PR00138; MATRIXIN.
DR SMART; SM00120; HK; 3.
DR SMART; SM00235; ZMNC; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR NON TER 1
DR NON TER 1
FT NON TER 1
SQ SEQUENCE 393 AA; 44093 MW; 148C78BAEBDA95 CRC64;

```

```

Query Match 8.7%; Score 189.5; DB 6; Length 393;
Best Local Similarity 26.4%; Pred. No. 1.1e-07;
Matches 63; Conservative 23; Mismatches 94; Indels 59; Gaps 8;

```

```

QY 25 PLPTSAHGNVAEGT-KDPDVTGRCSDGMSFPDATTLDNGTMLFKGEGFV-----KSHKMR 83
DB 201 PPPMSSPDGPAGPTTSVPEPETPATCDPALSFDVNS-SLRGELIFKDRHLMRKSLRTR 259
QY 84 E-----LISERWKNPSPVDAAFRQ--HNSVFLIKGRVWVYPPPEKKEGYPKLLQD--EPP 137
DB 260 EPGFILMCSFWSLSLSDGDAAYERTSKQIVLFFKNGQFWAMRGTRVQAGYPIGHTLGRF 319
QY 138 GLPSPLDAAVECHRGECQAGVLFPGHGHNGTGHGNSTHHGPEYMRCSPHLVLSALTS 197
DB 320 PYVKKIDAAV-----FD 331

```

```

QY 198 DNHGATYAAGSTHYWRDLTSRDGNH-SWP--IAHQWPGSPSAVDAAFSWEKLYLVQGT 253
DB 332 KKKKTYFVGDKYWRFEKRSMEBPGKQIAEDPGVDSTVDAAFEAFGFFYFENG 390

```

Search completed: December 16, 2003, 06:52:11  
 Job time : 179 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 15, 2003, 12:37:44 ; Search time 64 Seconds  
(without alignments)  
969.720 Million cell updates/sec

Title: US-09-900-448-2  
Perfect score: 2185  
Sequence: 1.MARVLGAPVALGLWLSLCSL.....NAKALPQPNVTSILGCTH 391

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2139.5	97.9	462	22	AA23933
2	456	20.9	83	21	AA600304
3	258	11.8	87	23	ABP31577
4	204	9.3	587	21	AA990502
5	204	9.3	587	21	AA98405
6	202.5	9.3	470	22	AAE10419
7	202.5	9.3	470	22	AA884614
8	202.5	9.3	470	22	AA874595
9	202.5	9.3	470	22	AA849982

10	202.5	9.3	470	23	AA91061	Human head and nec
11	202.5	9.3	470	24	ABP96800	Human COPD related
12	202.5	9.3	470	24	ABR48203	Human bladder canc
13	202.5	9.3	470	24	ABUS6421	Lung cancer-associ
14	202.5	9.3	470	24	ABUS6663	Lung cancer-associ
15	202.5	9.3	470	21	AA83772	Human cancer-associ
16	199	9.1	519	22	AAE10426	Human matrix metal
17	199	9.1	519	22	AA884619	Amino acid sequenc
18	199	9.1	605	21	AA990503	Human transmembran
19	199	9.1	605	24	AA019713	MT4-WMP catalytic
20	199	9.1	606	21	AA98404	Human MT4-WMP (2)
21	196	9.0	464	21	AAW52135	Rabbit matrix meta
22	195.5	8.9	477	23	AAU91062	Human head and nec
23	192.5	8.8	476	23	ABU03460	Angiogenesis-asso
24	189.5	8.7	475	23	ABU05574	Breast cancer-asso
25	189.5	8.7	475	24	ABUS6595	Lung cancer-associ
26	189.5	8.7	476	22	AAE10421	Human matrix metal
27	189.5	8.7	476	22	AA884612	Amino acid sequenc
28	189.5	8.7	476	24	ABR48147	Human bladder canc
29	189.5	8.7	476	24	ABP54455	Matrix metalloprot
30	183.5	8.4	444	22	AA65357	Human WMP-galt pol
31	183.5	8.4	467	22	AAE10416	Human matrix metal
32	183.5	8.4	467	22	AA884610	Amino acid sequenc
33	183.5	8.4	467	22	AAE10358	Human neutrophil c
34	181.5	8.3	469	22	AAE10415	Human matrix metal
35	181.5	8.3	469	22	AA884606	Amino acid sequenc
36	181.5	8.3	469	24	ABR48148	Human bladder canc
37	181.5	8.3	469	24	ABUS6596	Lung cancer-associ
38	181.5	8.3	469	24	ABUS6597	Lung cancer-associ
39	181.5	8.3	469	24	ABU07454	Protein different
40	181.5	8.3	469	24	ABP54454	Matrix metalloprot
41	181.5	8.3	492	13	AAE24863	Sequence of pre-pr
42	181.5	8.3	496	22	AA675509	Human colon cancer
43	180.5	8.3	457	10	AA93628	Sequence of human
44	180.5	8.3	469	8	AA970611	Sequence encoded b
45	179.5	8.2	479	23	ABP41861	Human ovarian anti

## ALIGNMENTS

RESULT 1	
AA23933	
ID	AA23933 standard; Protein: 462 AA.
AC	AA23933;
XX	
DT	12-OCT-2001 (first entry)
DE	Human EST encoded protein SEQ ID NO: 1458.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	W0200154477-A2.
XX	
PD	02-AUG-2001.
XX	
PR	25-JAN-2001; 2001WO-US02687.
XX	
PR	25-JUN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Gao Y, Dimaac RA, Zhang J, Werhman T;

```

XX  WPI; 2001-476164/51.
DR  N-PSDB; AAH98592.
XX
XX  Isolated polypeptide for treatment of diseases, diagnostics, raising
PT  antibodies and research use -
XX
XX  Claim 20; Page 1017-1018; 1275pp; English.
XX
XX  The present invention provides the protein and coding sequences of novel
CC  proteins from a variety of organisms, including human, dog, cat, horse,
CC  cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC  urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC  from the organism of interest. They can be used in diagnostics,
CC  forensics, gene mapping, identification of mutations, to assess
CC  biodiversity and for nutritional purposes. The present sequence is a
CC  protein of the invention.
XX
XX  Sequence 462 AA;
SQ
XX
XX  Query Match 97.9%; Score 2139.5; DB 22; Length 462;
XX  Best Local Similarity 84.6%; Pred. No. 8.9e-193;
XX  Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;
XX
XX  1 MARVLGAPVALGLMSLCMSLAIAITPLPTSAHGNAEGETKPPDVTTERCSGMSPDATT 60
XX  1 MARVLGAPVALGLMSLCMSLAIAITPLPTSAHGNAEGETKPPDVTTERCSGMSPDATT 60
XX
XX  61 LDNGTMTLFFKGEFVKSHKMDRELISERKCNFPSPVDAAFROGHNSVFLIKDKYVVP 120
XX  61 LDNGTMTLFFKGEFVKSHKMDRELISERKCNFPSPVDAAFROGHNSVFLIKDKYVVP 120
XX
XX  121 PEKKEGYPLTLDPEFGIPSPIDAAYECHRGEOAEVGFQ----- 163
XX  121 PEKKEGYPLTLDPEFGIPSPIDAAYECHRGEOAEVGFQ----- 163
XX
XX  164 -----GHGHRN 169
XX  164 -----GHGHRN 169
XX
XX  181 SWPVAVNCSSALMLWLGRIYCFQGNQFLRPDPVARGVPRPRVROVDFMPCPGRHGHRN 240
XX  181 SWPVAVNCSSALMLWLGRIYCFQGNQFLRPDPVARGVPRPRVROVDFMPCPGRHGHRN 240
XX
XX  170 GTCHGNSHTHGPYMRCSPLVLTALTSDNHGATVAFSGTHYWRDLTSDRGMSWPIAQ 229
XX  170 GTCHGNSHTHGPYMRCSPLVLTALTSDNHGATVAFSGTHYWRDLTSDRGMSWPIAQ 229
XX
XX  241 GTCHGNSHTHGPYMRCSPLVLTALTSDNHGATVAFSGTHYWRDLTSDRGMSWPIAQ 300
XX  241 GTCHGNSHTHGPYMRCSPLVLTALTSDNHGATVAFSGTHYWRDLTSDRGMSWPIAQ 300
XX
XX  230 WPGSPSAVDAAFSWEKLYVVGQVYVFLTKGYTLVSGYPRLEKEVGTPIGILLDSV 289
XX  230 WPGSPSAVDAAFSWEKLYVVGQVYVFLTKGYTLVSGYPRLEKEVGTPIGILLDSV 289
XX
XX  301 WPGSPSAVDAAFSWEKLYVVGQVYVFLTKGYTLVSGYPRLEKEVGTPIGILLDSV 360
XX  301 WPGSPSAVDAAFSWEKLYVVGQVYVFLTKGYTLVSGYPRLEKEVGTPIGILLDSV 360
XX
XX  290 DAAFIKPGSSRLHIMAGRLIMWLDLKSGAQTWTELPWPEKVDGALCMEKSLGPNSSCA 349
XX  290 DAAFIKPGSSRLHIMAGRLIMWLDLKSGAQTWTELPWPEKVDGALCMEKSLGPNSSCA 349
XX
XX  361 DAAFIKPGSSRLHIMAGRLIMWLDLKSGAQTWTELPWPEKVDGALCMEKSLGPNSSCA 420
XX  361 DAAFIKPGSSRLHIMAGRLIMWLDLKSGAQTWTELPWPEKVDGALCMEKSLGPNSSCA 420
XX
XX  350 NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPONVTSILGCTH 391
XX  350 NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPONVTSILGCTH 391
XX
XX  421 NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPONVTSILGCTH 462
XX  421 NGPGLYLHIGPNLYCYSDVEKLNAAKALPOPONVTSILGCTH 462
XX
XX  RESULT 2
XX  AAG00304
XX  ID AAG00304 standard; Protein; 83 AA.
XX
XX  AAG00304;
XX
XX  06-OCT-2000 (first entry)
XX
XX  Human secreted protein, SEQ ID NO: 4385.
XX
XX  Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX  gene therapy; chromosome mapping.
XX
XX  Homo sapiens.
XX
XX  EP1033401-A2.

```

```

XX  06-SEP-2000.
XX
XX  21-FEB-2000; 2000EP-020610.
XX
XX  26-FEB-1999; 99US-0122487.
XX
XX  (GEST ) GENSET.
XX
XX  Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX  WPI; 2000-500381/45.
XX
XX  N-PSDB; AAC00310.
XX
XX  New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT  obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT  diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX  Claim 13; SEQ ID 4385; 71pp + CD-ROM; English.
XX
XX
XX  The present sequence is a polypeptide encoded by one of a large number
CC  of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC  were prepared from total human RNA or polyA+ RNAs derived from 30
CC  different tissues. EST sequences usually correspond mainly to the 3'
CC  untranslated region (UTR) of the mRNA because they are often obtained
CC  from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC  isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC  those cases where longer cDNA sequences have been obtained, the full 5'
CC  UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC  ends and can therefore be used to obtain full length cDNAs and genomic
CC  DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC  chromosome mapping procedures. They are used to obtain upstream
CC  regulatory sequences and to design expression and secretion vectors.
XX
XX  Sequence 83 AA;
SQ
XX
XX  Query Match 20.9%; Score 456; DB 21; Length 83;
XX  Best Local Similarity 98.8%; Pred. No. 3.9e-35;
XX  Matches 82; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX  1 MARVLGAPVALGLMSLCMSLAIAITPLPTSAHGNAEGETKPPDVTTERCSGMSPDATT 60
XX  1 MARVLGAPVALGLMSLCMSLAIAITPLPTSAHGNAEGETKPPDVTTERCSGMSPDATT 60
XX
XX  61 LDNGTMTLFFKGEFVKSHKMDR 83
XX  61 LDNGTMTLFFKGEFVKSHKMDR 83
XX
XX  RESULT 3
XX  ABP31577
XX  ID ABP31577 standard; Protein; 87 AA.
XX
XX  ABP31577;
XX
XX  09-JUL-2002 (first entry)
XX
XX  Human glycoprotein-like ORF550 protein, SEQ ID NO:1100.
XX
XX
XX  Human; ORF; open reading frame; ORF; drug screening; diagnosis;
XX  disease monitoring; cytokine; cell proliferation; cell differentiation;
XX  immune modulation; haematopoiesis regulation; tissue growth;
XX  angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;
XX  thrombolytic; tumour inhibition; bodily characteristics; fertility;
XX  behaviour; cancer; proliferative disorder; neurological disorder;
XX  cardiovascular disease; immune system disorder; organ transplantation;
XX  tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
XX  hypothyroidism; cholesterol ester storage disease; infection; vulvareary;
XX  vasotropic; antipsoriatic; antidiabetic; cyostatic; nootropic;
XX  neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
XX  cardiatic; hypotensive; antihypertensive; antiinflammatory; immunomodulator;
XX  dermatological; analgesic; virucide; antibacterial; fungicide.

```

**OS** Homo sapiens.  
**XX**

**PN** WO200190366-A2.  
**XX**

**XX** 29-NOV-2001.  
**XX**

**PF** 24-MAY-2001; 2001WO-US17076.  
**XX**

**PR** 24-MAY-2000; 2000US-206690P.  
**XX**

**PA** (CURA-) CURAGEN CORP.  
**PI** Leach MD, Shinkets RA;  
**DR** WPI; 2002-106200/14.  
**N-PSDB;** ABN75603.

Novel human polypeptides and polymucleotides useful for diagnosing,  
preventing and treating cardiovascular disease, neurodegenerative,  
hyperproliferative disorders and disorders related to organ  
transplantation -  
PT  
PS Claim 10; Page 541-542; 2508pp; English.  
XX

Sequences ABP1028-ABP35561 represent 4534 novel human proteins  
designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
ABN79587 represent cDNAs encoding them. The invention also encompasses  
polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
referred to as ORFX) proteins, polymucleotides at least 85% identical to  
the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
polymucleotides, the recombinant production of detecting ORFX proteins, antibodies  
specific for ORFX proteins, methods of screening for modulators of ORFX expression and  
polypeptides, methods of screening for modulators of ORFX expression or  
activity, and methods of screening individuals for a predisposition to an  
ORFX-associated disorder. The ORFX proteins of the invention have a wide  
range of biological activities, such as cytokine, cell proliferation,  
cell differentiation, immune modulation, haematopoiesis regulation,  
tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
chemokinetic activity, haemostatic activity, thrombolytic activity,  
receptor/ligand, anti-inflammatory activity, tumour inhibition activity,  
CC and antifibrotic activity, and may also be involved in the determination  
of bodily characteristics, fertility and behaviour. ORFX proteins,  
nucleic acids and antibodies may be used in the treatment of cancers,  
other proliferative disorders such as psoriasis and benign tumours,  
neurological disorders such as epilepsy and Alzheimer's disease,  
cardiovascular diseases, immune system disorders, disorders related to  
organ transplantation, disorders of tissue growth and regeneration,  
diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
storage disease, and infectious diseases caused by viral, bacterial,  
fungal and other pathogens. ORFX nucleic acids may also be used as a  
source of primers and probes, in the detection of ORFX genomic sequences  
or transcripts, in the identification and cloning of homologous  
sequences, in genetic diagnosis, and in forensic biology. The ORFX  
nucleic acids may additionally be used to produce transgenic animals  
which may be useful for studying the function and/or activity of ORFX  
protein, and in drug screening. The ORFX proteins may also be used as  
immunogens to generate specific antibodies, which are useful in the  
diagnosis, treatment and monitoring of ORFX-associated diseases.

**SQ** Sequence 87 AA;  
**XX**

Query Match 11.8%; Score 258; DB 23; Length 87;  
Best Local Similarity 86.5%; Pred. No. 2e-16;  
Matches 45; Conservative 3; Mismatches 4; Indels 0; Gaps 0

114 DKWWYPPPEKKKGYPKLIDFFPGIPSLDAAVCHRGCECAEGYLFPQGH 155  
|||||:  
S DKWVVYPPEKKENGYPKLFGEBFPGIYPDPDAAVECHRGCESGVLFPQGN 56

RESULT 4  
AAV90502  
ID AAV90502 standard; Protein: 587 AA.

```

XX AC AAY90502;
XX DT 15-AUG-2000 (first entry)
XX DE Murine transmembrane matrix metalloprotease MT4-MMP (2).
XX XX
XX Matrix metalloprotease; MT4-MMP (2); transmembrane; mouse; antibody;
XX drug screening; diagnosis; arthrosis deformans; rheumatoid arthritis;
XX asthma; autoimmune disease; atopic dermatitis; skin disorder;
XX inflammatory disorder; arteriosclerosis; cancer; tissue damage;
XX OS Mus sp.
XX PN WO20001805-A1.
XX XX
XX 06-APR-2000.
XX PD
XX PF 29-SEP-1999; 99WO-JP05350.
XX PR 29-SEP-1998; 98JP-0291501.
XX PR 29-SEP-1998; 98JP-0291503.
XX PA (KIOWA) KIOWA HAKKO KOGYO KK.
XX PI Hanai N, Furuya A;
XX DR MPI: 2000-293115/25.
XX DR N-PSDB; AAA14339.
XX PT Antibody immunospecific for transmembrane matrix metalloprotease
XX PT polypeptide, useful in screening of inhibitors and activators as drugs
XX PT for, and for diagnosis of e.g. arthrosis deformans and asthma
XX PS Claim 1; Page 86-90; 152pp; Japanese.
XX XX
XX This sequence represents a novel murine transmembrane matrix
XX CC metalloprotease MT4-MMP (2). MT4-MMP (2) has physiological activity
XX CC different to that of MT4-MMP. The invention relates to antibodies
XX CC against both murine and human MT4-MMP (2) (AAY90502 and AAY90503) and to
XX CC antibodies against a novel murine and human transmembrane matrix
XX CC metalloprotease MT5-MMP (AAY90504 and AAY90505). The antibodies are
XX CC useful for the diagnosis and screening of inhibitors and activators
XX CC useful for treating or preventing a wide variety of medical conditions,
XX CC such as arthrosis deformans, rheumatoid arthritis, asthma, autoimmune
XX CC diseases and atopic dermatitis. The anti-MT5-MMP antibodies may
XX CC additionally be used to diagnose and screen therapeutic or prophylactic
XX CC agents for conditions affecting the brain, such as cerebral stroke and
XX CC Alzheimer's disease. The antibodies of the invention may also be used for
XX CC the prevention, diagnosis and treatment of psoriasis, contact
XX CC dermatitis, hair loss, ischaemic diseases, immune reaction accompanying
XX CC organ transplant, hepatitis, nephritis, pancreatitis, arteriosclerosis,
XX CC leukaemia, malignant tumours, wounds, corneal ulcers, tissue damage or
XX CC inflammation accompanying leukocytic infiltration, dementia, multiple
XX CC sclerosis, Parkinson's disease or brain tumour.
XX SQ
XX Sequence 587 AA;
QY Query Match 9.3%; Score 204; DB 21; Length 587;
QY Best Local Similarity 24.4%; Pred. No. 3.9e-10;
QY Matches 103; Conservative 34; Mismatches 127; Indels 158; Gaps 25
DB 5 LGAPNALGL-----WSLCSLAIALPL-----PPTSHGNAVAEETKPPDDY 46
DB 273 VGDPFRYGLPEDRVRVQLGVNBSVSPYQLDTPBEBEPPLLEPPNNNSSTPPQXDV 332
QY 47 TERCSDDGSPFATTLDDNGTALFPKGEFVMSKHKVRELISSR-----WKNFP---SP 96
QY 333 PHRCGA--HFPAYVA-QIRGEAFPFKGYKFWLTLR--DRILVELQLPQOMHFRKGLPLHIDS 388
QY 97 VDAAF-ROGHNSVFLIKGDKVWVYVPEKKEKGYPKLQDEPPGLIPDAAVECHRGECQ 155

```

```

Db      389 VDAVVERTSDHKIVFPGDRYVWFKDNVVEGYRPRVSD-FSLPPGIDA----- 437
QY      156 AEGVLFPGQHGHNCHNGSTHNGPVMCSPHLVLSALTSNHNATYAFSGTHWRD 215
Db      438 ----VFSWANDR-----TYFFKQQLWRYD 459
QY      216 --TSR--DGWHS-WPIAHQWPGPSAVDAAFSWE-KLYLVQGTQVYVFLTKGGYTLVSG 269
Db      460 DHTRRMDPGYPAQGPL---WRGVPSMLDDMRMSDGSFYFRG-QEYWKVLDSLELAAPG 515
QY      270 YPRLEKE---VGTPIGILLSDVDAFICPGSSRLHIMAGRRLWLDLKSGAQTWTELP 326
Db      516 YPQSTARDMLVCGEP-----LADAEVGPQPQ-----GR-----SGAQ----- 548
QY      327 WPEHKVDG-ALCMKESLGRNSCSANGPGLVLIHGPNLYCYSDVEKLNAAKALPQPNVTS 385
Db      549 ----DGLAVCS-----CTSDAHL-----ALPSLLLLTP 573
QY      386 LL 387
Db      574 LL 575

RESULT 5
AAV88405
ID      AAV88405 standard; Protein; 587 AA.
XX
AC      AAV88405;
XX
DT      28-JUL-2000 (first entry)
XX
DE      Mouse MT4-MMP (2) matrix metalloproteinase amino acid sequence.
XX
KW      Matrix metalloproteinase; MT4-MMP; treatment; arthritis deformans;
KW      Chronic Rheumatoid arthritis; asthma; autoimmune disease; hepatitis;
KW      atrophic dermatitis; psoriasis; contact dermatitis; hair loss; nephritis;
KW      ischaemic disease; pancreatitis; arteriosclerosis; leukaemia; tumour;
KW      wound; corneal ulcer; tissue damage; inflammation; cerebral apoptosis;
KW      Alzheimer's disease; dementia; multiple sclerosis; Parkinson's disease;
KW      gene therapy.
XX
OS      Mus sp.
XX
PN      WO200018900-A1.
XX
PD      06-APR-2000.
XX
PF      29-SEP-1999; 99WO-JP05349.
XX
PR      29-SEP-1998; 98JP-0276258.
PR      29-SEP-1998; 98JP-0291505.
XX
PA      (SEIKI) SEIKI M.
XX
PI      Seiki M;
XX
DR      WPI: 2000-293141/25.
XX      N-PSDB; AAA13378.
XX
PT      Physiologically-active transmembrane matrix metalloproteinase
PT      polypeptide, useful in screening inhibitors and activators for treating
PT      e.g. arthritis deformans, asthma and cancers
XX
PS      Claim 11; Page 59-66; 113pp; Japanese.
XX
CC      This sequence represents the murine matrix metalloproteinase (MT4-MMP 2)
CC      amino acid sequence. The invention relates to a mammalian transmembrane
CC      MMP polypeptide, its variants and the DNA sequences encoding them. Also
CC      included in the invention is a vector comprising the MMP DNA molecules,
CC      and cells transformed using the vector. The protein, DNA and antibodies
CC      directed against the polypeptide are useful in screening inhibitors and
CC      activators for use in treating arthritis deformans, chronic rheumatoid
CC      arthritis, asthma, autoimmune diseases, atrophic dermatitis, psoriasis,

```

```

CC      contact dermatitis, hair loss, ischaemic diseases, immune reaction
CC      accompanying organ transplant, hepatitis, nephritis, pancreatitis,
CC      arteriosclerosis, leukaemia, malignant tumours, wounds, corneal ulcers,
CC      tissue damage or inflammation accompanying leukocytic infiltration,
CC      together with brain disorders during cerebral apoptosis, Alzheimer's
CC      disease, dementia, multiple sclerosis, Parkinson's disease or brain
CC      tumours. The DNA can also be incorporated into a suitable vector for use
CC      in gene therapy.
XX
SQ      Sequence 587 AA;
XX
Query Match 9.3%; Score 204; DB 21; Length 587;
Best Local Similarity 24.4%; Pred. No. 3.9e-10;
Matches 103; Conservative 34; Mismatches 127; Indels 158; Gaps 25;

QY      5 LGAPVALGL-----WELCMSLATATPL-----PPRSAGNVABEGTKRDPV 46
Db      273 VGDPRVYGLPEYEDRYVWQLYGVRESVSPYALDTPPEPEPPLPPEPPNRRSTPPQKV 332
QY      47 TERCSQWSPDATTLDDNGTMLPFKGFVWKSHKMDRELISER-----WKNFP--SP 96
Db      333 PIRCTA--HPDAVA-QIRGEAFFKGYRRLTR-DRLVSLQPAQMHKPLHLDS 388
QY      97 VDAAF-ROGHSVFLIKGDKRWVYVPEKKEKGYPLLODEPPIPSPLDAVCHRGECQ 155
Db      389 VDAVVERTSDHKIVFPGDRYVWFKDNVVEGYRPRVSD-FSLPPGIDA----- 437
QY      156 AEGVLFPGQHGHNCHNGSTHNGPVMCSPHLVLSALTSNHNATYAFSGTHWRD 215
Db      438 ----VFSWANDR-----TYFFKQQLWRYD 459
QY      216 --TSR--DGWHS-WPIAHQWPGPSAVDAAFSWE-KLYLVQGTQVYVFLTKGGYTLVSG 269
Db      460 DHTRRMDPGYPAQGPL---WRGVPSMLDDMRMSDGSFYFRG-QEYWKVLDSLELAAPG 515
QY      270 YPRLEKE---VGTPIGILLSDVDAFICPGSSRLHIMAGRRLWLDLKSGAQTWTELP 326
Db      516 YPQSTARDMLVCGEP-----LADAEVGPQPQ-----GR-----SGAQ----- 548
QY      327 WPEHKVDG-ALCMKESLGRNSCSANGPGLVLIHGPNLYCYSDVEKLNAAKALPQPNVTS 385
Db      549 ----DGLAVCS-----CTSDAHL-----ALPSLLLLTP 573
QY      386 LL 387
Db      574 LL 575

RESULT 6
AAE10419
ID      AAE10419 standard; Protein; 470 AA.
XX
AC      AAE10419;
XX
DT      10-DEC-2001 (first entry)
XX
DE      Human matrix metalloproteinase-12 (MMP-12) protein.
XX
KW      Human; matrix metalloproteinase; MMP-12; hair growth; antisense therapy;
KW      endopeptidase; skin cell; breast cancer; hair follicle; chromosome 11q22.
XX
OS      Homo sapiens.
XX
FH      Key
FH      Peptide 1..16 Location/Qualifiers
FT      Protein /label= Signal_peptide 17..470
FT      Domain /label= Mature_MMP_12_protein 90..96
FT      Domain /label= Cysteine_switch_domain 161..185
FT      Domain /note= "Zinc and calcium binding domain"
XX

```



PN WO200166766-A2.  
XX 13-SEP-2001.  
XX 06-MAR-2001; 2001WO-US07167.  
XX 06-MAR-2001; 2000US-0187196.  
XX (DARW-) DARWIN MOLECULAR CORP.  
XX (SCHA/) SCHATZMAN R.  
XX Fajardo M, Wang K, Smith R, Moss P;  
XX WPI; 2001-582276/65.  
XX Novel isolated matrix metalloproteinase-25 nucleic acid molecule and  
PT proteins encoded by them whose inhibition is useful for modulation of  
PT hair growth in mammals -  
XX  
XX Example 2; Fig 3; 119pp; English.  
XX The present sequence is human matrix metalloproteinase (MMP)-12 protein  
CC used in the exemplification of the invention. MMP-25 DNA is located  
CC on chromosome 11q22. Matrix metalloproteinases are a family of zinc  
CC dependent endopeptidases that function extracellularly to degrade  
CC proteins typically found in the extracellular matrix. MMP-25 is expressed  
CC in skin cells of mammals, particularly in breast cells and hair  
CC follicles. MMP-25 DNA is useful for identifying a nucleic acid molecule  
CC encoding all or part of MMP by hybridizing MMP-25 to a nucleic acid  
CC sample and identifying a sequence that hybridizes in the nucleic acid  
CC acid sample. The identification step involves performing polymerase  
CC chain reaction (PCR) to amplify the hybridizing sequence. MMP-25  
CC antibody is useful for identifying type 25 MMP. MMP-25 protein  
CC inhibitors may be used to modulate hair growth and breast cancer in  
CC a mammal.  
XX  
SQ Sequence 470 AA;  
XX  
Query Match 9.3%; Score 202.5; DB 22; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
QY 30 SAHGNAVEGKTPDPVTE--RCSGWSFDATTLDDNGTMLPFKGEFVW---KSHKMDR 83  
DB 260 SLVGDPEKNGRLPRPNDSBPALCDPNLSFDAVTVGN-KIFFKDRFWMKVSERPTSV 318  
QY 84 ELISERKKNFSPVDAFR-QGHSVFLIKGDVWVYPPPEKKGYPKLQD-BFPGIPS 141  
DB 319 NLISLWPTLPSCGIEAAVEIARNOVFLFKDDKYWLISNLRPENYPKSIHSFGFPNFK 378  
QY 142 PLDAAVECHRGECQAEGLFFQGHGHRNGTGHGNSTHHGPEYMKCPHLVLSALTSDNHG 201  
DB 379 KIDAAV-----FNPRFYR----- 391  
QY 202 ATYAFSGTHWRLDTSRD---GHSWPIAHQWPGPSAVDAFSEWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNOYWRVDERROMDPPGYPKLITKNFGQIGPK-IDAVFYSKNKXYFFQGSNQP 449  
QY 257 VF 258  
DB 450 EY 451  
RESULT 7  
AAB84614  
ID AAB84614 standard; Protein; 470 AA.  
XX AAB84614;  
XX  
XX 05-SEP-2001 (first entry)  
XX  
XX Amino acid sequence of matrix metalloproteinase-12.

KW Growth factor; protein inhibitor; protease; damaged tissue;  
KW platelet-derived growth factor; PDGF; fibroblast growth factor; FGF;  
KW connective tissue derived growth factor; CTGF; chrysalin; VEGF;  
KW keratinocyte-derived growth factor; KGF; epidermal growth factor; EGF;  
KW transforming growth factor-beta; TGF-beta; matrix metalloproteinase; MMP;  
KW granulocyte macrophage colony stimulating factor; GM-CSF; uPA;  
KW vascular endothelial growth factor; urokinase plasminogen activator;  
KW dermal ulcer; wound.  
XX  
XX Homo sapiens.  
XX  
XX WO200149309-A2.  
XX 12-JUL-2001.  
XX  
XX 21-DEC-2000; 2000WO-IB01935.  
XX  
XX 29-DEC-1999; 99GB-0030768.  
XX  
XX (PRIZ ) PRIZER LTD.  
XX (PRIZ ) PRIZER INC.  
XX  
XX Davies MJ, Huggins JP, McIntosh FS, Occlleston NL;  
XX WPI; 2001-418351/44.  
XX N-PSDB; AAB28229.  
XX  
XX Composition for the treatment of damaged tissue i.e. chronic wounds and  
PT dermal ulcers comprises an inhibitor agent i.e. a protease and a growth  
PT factor -  
XX  
XX Disclosure; Page 560; 572pp; English.  
XX  
XX The specification describes a pharmaceutical composition, comprising  
CC a growth factor, an inhibitor agent, i.e. a protease. The inhibitor  
CC agent inhibits the action of at least one specific adverse protein,  
CC i.e. a protease, that is upregulated in a damaged tissue such as a  
CC wound environment. Growth factors which are included in the composition  
CC of the invention are platelet-derived growth factor (PDGF), fibroblast  
CC growth factor (FGF), connective tissue derived growth factor (CTGF),  
CC keratinocyte-derived growth factor (KGF), transforming growth  
CC factor-beta (TGF-beta), granulocyte macrophage colony stimulating factor  
CC (GM-CSF), epidermal growth factor (EGF), vascular endothelial growth  
CC factor (VEGF), and chrysalin. Inhibitors which are included in the  
CC composition of the invention include inhibitors of urokinase-type  
CC plasminogen activator (uPA) and matrix metalloproteinase (MMP). The  
CC composition is useful for the treatment of chronic damaged tissue, i.e.  
CC wounds and dermal ulcers. The present sequence represents a human MMP-12,  
CC and is used to produce the composition of the invention.  
XX  
SQ Sequence 470 AA;  
XX  
Query Match 9.3%; Score 202.5; DB 22; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
QY 30 SAHGNAVEGKTPDPVTE--RCSGWSFDATTLDDNGTMLPFKGEFVW---KSHKMDR 83  
DB 260 SLVGDPEKNGRLPRPNDSBPALCDPNLSFDAVTVGN-KIFFKDRFWMKVSERPTSV 318  
QY 84 ELISERKKNFSPVDAFR-QGHSVFLIKGDVWVYPPPEKKGYPKLQD-BFPGIPS 141  
DB 319 NLISLWPTLPSCGIEAAVEIARNOVFLFKDDKYWLISNLRPENYPKSIHSFGFPNFK 378  
QY 142 PLDAAVECHRGECQAEGLFFQGHGHRNGTGHGNSTHHGPEYMKCPHLVLSALTSDNHG 201  
DB 379 KIDAAV-----FNPRFYR----- 391  
QY 202 ATYAFSGTHWRLDTSRD---GHSWPIAHQWPGPSAVDAFSEWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNOYWRVDERROMDPPGYPKLITKNFGQIGPK-IDAVFYSKNKXYFFQGSNQP 449  
QY 257 VF 258

Db 450 EY 451

## RESULT 8

ID AAB74595 standard; Protein; 470 AA.

AC AAB74595;

DT 11-JUN-2001 (first entry)

DE Human macrophage metalloelastase HME.

KM Human; HME: macrophage metalloelastase; elastolysis; pulmonary emphysema; inflammatory disease.

OS Homo sapiens.

PN US6204043-B1.

PD 20-MAR-2001.

PF 01-MAR-1995; 95US-0396988.

PR 28-MAY-1993; 93US-0068392.

PA (UNIW ) UNIV WASHINGTON.

PI Shapito SD;

DR WPI; 2001-289518/30.

DR N-PSDB; AAF81624.

PT New human macrophage metalloelastase, useful in the normal embryonic development, growth, tissue remodeling and tissue repair, particularly in studying the pathogenesis of pulmonary emphysema -

PS Claim 1; Fig 5; 20pp; English.

CC The present invention provides the protein and coding sequences of human macrophage metalloelastase (HME). The protein has elastolytic activity.

CC HME has a role in tissue remodeling and repair associated with development and inflammation, and abnormal expression can result in tumour invasiveness, arthritis and atherosclerosis. It is also thought to contribute to the pathogenesis of pulmonary emphysema and other inflammatory destructive diseases. The present sequence is the HME protein.

CC Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 22; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10; Mismatches 90; Indels 63; Gaps 10;

Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGETKPDVTE--RCSGWSFDATLTDNGTMLFFKGEFFW---KSHKMDR 83

DB 260 SLVGDPKRNQLRPNPNSPALCDPMLSFDAVTVGN-KIFFKDRFVLKVSERPKTSV 318

QY 84 ELISEERKKNPPSPVDAFR-QGHNSVFLIKGDKVWYRPEKKEGYPKLQD-EFGGIPS 141

DB 319 NLISLSMPTLPSCIEAAYEIEARNQVFLPKDKYWLISNLRPEPNYPKSIHSPGFNPFVK 378

QY 142 PLDAAYECHRGEQAGVLFQGHGHRNGTGHNSTHHGPEYVRCSPHLVLSALTSNMG 201

DB 379 KIDAAV-----FNPRFYR----- 391

QY 202 ATYAFSGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAFASWEK-LYLVQGTQVY 256

DB 392 -TYFFVNDQWRYRDERQMMDPGYPKLITKNFGGIGPK-IDAVFYSGKNKYVYFFQGSNPF 449

QY 257 VF 258

Db 450 EY 451

## RESULT 9

ID AAB49982 standard; Protein; 470 AA.

AC AAB49982;

DT 13-MAR-2001 (first entry)

DE Human macrophage metalloelastase.

KM Human; macrophage metalloproteinase; elastin; matrix degrading enzyme; emphysema.

OS Homo sapiens.

PN US6150152-A.

PD 21-NOV-2000.

PF 28-MAY-1993; 93US-0068392.

PR 28-MAY-1993; 93US-0068392.

PA (UNIW ) UNIV WASHINGTON.

PI Shapito SD;

DR WPI; 2001-049090/06.

DR N-PSDB; AAC64994.

PT Novel human macrophage metalloelastase polynucleotides and polypeptides useful for measuring elastin degradation -

PS Claim 1; Fig 5; 20pp; English.

CC The present invention provides the coding and protein sequences for the human macrophage metalloelastase. This is a matrix degrading metalloproteinase which has the ability to degrade elastin, and can be used to measure elastin degradation.

CC Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 22; Length 470;

Best Local Similarity 27.3%; Pred. No. 3.9e-10; Mismatches 90; Indels 63; Gaps 10;

Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGETKPDVTE--RCSGWSFDATLTDNGTMLFFKGEFFW---KSHKMDR 83

DB 260 SLVGDPKRNQLRPNPNSPALCDPMLSFDAVTVGN-KIFFKDRFVLKVSERPKTSV 318

QY 84 ELISEERKKNPPSPVDAFR-QGHNSVFLIKGDKVWYRPEKKEGYPKLQD-EFGGIPS 141

DB 319 NLISLSMPTLPSCIEAAYEIEARNQVFLPKDKYWLISNLRPEPNYPKSIHSPGFNPFVK 378

QY 142 PLDAAYECHRGEQAGVLFQGHGHRNGTGHNSTHHGPEYVRCSPHLVLSALTSNMG 201

DB 379 KIDAAV-----FNPRFYR----- 391

QY 202 ATYAFSGTHYRLDTSRD---GMHSWPIAHQWPGPSAVDAFASWEK-LYLVQGTQVY 256

DB 392 -TYFFVNDQWRYRDERQMMDPGYPKLITKNFGGIGPK-IDAVFYSGKNKYVYFFQGSNPF 449

QY 257 VF 258

DB 450 EY 451

QY 257 VF 258

ID AAB49982 standard; Protein; 470 AA.

XX AAU91061;  
AC 05-JUN-2002 (first entry)  
XX Human head and neck tumour protein, SEQ ID No 219.  
XX Human head and neck cancer; tumour; cytostatic; immunogenic; vaccine.  
XX Homo sapiens.  
XX WO200212329-A2.  
XX 14-FEB-2002.  
XX 01-AUG-2001; 2001WO-US24226.  
XX 03-AUG-2000; 2000US-223281P.  
XX 16-NOV-2000; 2000US-249933P.  
XX (CORI-) CORIXA CORP.  
XX Wang T, Fan L;  
XX WPI; 2002-257467/30.  
XX N-PSDB; ABK54023.  
XX Novel polynucleotide encoding head and neck tumour polypeptides, useful  
XX in pharmaceutical compositions, e.g. vaccines, for treating head and  
XX neck cancers -  
XX Claim 2; Page 181-182; 200P; English.  
XX The invention relates to an isolated polynucleotide (I) comprising  
XX sequences selected from 273 sequences fully defined in the specification.  
XX (I), including its encoded polypeptide (II), an antibody binding to (II),  
XX a fusion protein comprising (II) and a T-cell population stimulated by  
XX (I) or (II) are useful for stimulating an immune response in a patient  
XX and treating head and neck cancer in a patient. An oligonucleotide (III)  
XX that hybridises to (I) is useful for determining the presence of cancer  
XX in a patient, by obtaining a biological sample from the patient,  
XX contacting the sample with (III), detecting in the sample an amount of a  
XX polynucleotide that hybridises to the oligonucleotide, and comparing the  
XX amount of polynucleotide that hybridises to the oligonucleotide to a  
XX predetermined cut-off value. (I) and (II) are useful in pharmaceutical  
XX compositions, e.g. vaccines, and other compositions for the diagnosis and  
XX treatment of head and neck cancer. AAU91061-AAU91062 represent human head  
XX and neck cancer protein sequences of the invention.  
XX Sequence 470 AA;  
SQ

Query Match 9.3%; Score 202.5; DB 23; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
OY 30 SAHGNVAGETKPPDYTE--RCSGWSFDTATLDNGTMLFFKGEFVW---KSHKMDR 83  
DB 260 SLYGDPRKNQRLPNDNSEPALCDPNLSFDAVTVGN-KIFFFDKRFMLKVSEKPTSV 318  
OY 84 ELISERKKNPSPVDAAPR-QGHSVFLIKGDKVWVPEKKEKGYKLLD-BEPGIPS 141  
DB 319 NLISLMPFTLPSGEIAVEIARNGVFLFKDQKWLISNLRPEPNYPKSHSPGPNPVK 378  
OY 142 PLDAAVECHRGCECAEGVLFQGHGHRNGTGHGNSTHHGPEYMRCSPLVLVSLATSDNHG 201  
DB 379 KIDAAV-----FNPRFYR----- 391  
OY 202 ATYAFSGTHYRLDTSRD---GHSWPIAHOWPGSAPVDAAPSMEEK-LYLVGQTQVY 256  
DB 392 -TYFFVNDQYWRBYDERROMDPBGYPKLLTKNFGIGPK-IDAVFYSKKYYTFFQGSNOF 449  
OY 257 VF 258

DB 450 EY 451  
RESULT 11  
ABP96800  
ID ABP96800 standard; Protein, 470 AA.  
XX ABP96800;  
XX 05-JUN-2003 (first entry)  
XX Human COPD related protein SEQ ID NO:50.  
XX Human, chronic obstructive pulmonary disease; COPD; chronic lung disease.  
XX Homo sapiens.  
XX WO200297127-A2.  
XX 05-DEC-2002.  
XX 28-MAY-2002; 2002WO-EP05835.  
XX 31-MAY-2001; 2001GB-0013266.  
XX (FARB ) BAYER AG.  
XX Oellers N, Gehrmann M, Kallabis H, Hall R, Schulze T, Kroegel C;  
XX WPI; 2003-140492/13.  
XX N-PSDB; ACC46771.  
XX Predicting, diagnosing or prognosing chronic lung disease, by detecting  
XX a chronic obstructive pulmonary disease (COPD) gene in a biological  
XX sample -  
XX Claim 8; Page 153-155; 214P; English.  
XX The present invention describes a method for predicting, diagnosing or  
XX prognosing chronic lung disease by detecting a chronic obstructive  
XX pulmonary disease (COPD) gene related polynucleotide (see ACC46750 to  
XX ACC46777, which encode the COPD related proteins in ABP96779 to  
XX ABP96806). The method is useful for predicting, diagnosing or prognosing  
XX chronic lung disease in a biological sample. The COPD genes and proteins  
XX encoded by them from the present invention (I) can be used for treating  
XX or preventing chronic lung disease in a mammal. (I) can be used in an  
XX animal model for determining the efficacy, toxicity, or side effects of  
XX treatment with (I), and determining the mechanism of action of (I).  
XX ACC46778 to ACC46903 represent COPD related PCR primers and probes used  
XX in an example from the present invention.  
XX Sequence 470 AA;  
SQ

Query Match 9.3%; Score 202.5; DB 24; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
OY 30 SAHGNVAGETKPPDYTE--RCSGWSFDTATLDNGTMLFFKGEFVW---KSHKMDR 83  
DB 260 SLYGDPRKNQRLPNDNSEPALCDPNLSFDAVTVGN-KIFFFDKRFMLKVSEKPTSV 318  
OY 84 ELISERKKNPSPVDAAPR-QGHSVFLIKGDKVWVPEKKEKGYKLLD-BEPGIPS 141  
DB 319 NLISLMPFTLPSGEIAVEIARNGVFLFKDQKWLISNLRPEPNYPKSHSPGPNPVK 378  
OY 142 PLDAAVECHRGCECAEGVLFQGHGHRNGTGHGNSTHHGPEYMRCSPLVLVSLATSDNHG 201  
DB 379 KIDAAV-----FNPRFYR----- 391  
OY 202 ATYAFSGTHYRLDTSRD---GHSWPIAHOWPGSAPVDAAPSMEEK-LYLVGQTQVY 256  
DB 392 -TYFFVNDQYWRBYDERROMDPBGYPKLLTKNFGIGPK-IDAVFYSKKYYTFFQGSNOF 449

QY 257 VF 258  
DB 450 EY 451

## RESULT 12

ABR48203  
ID ABR48203 standard; Protein; 470 AA.

XX ABR48203;

DT 12-JUN-2003 (first entry)

DE Human bladder cancer associated protein sequence SEQ ID NO:125.

KM Human; bladder cancer; cytostatic; gene therapy; vaccine.

OS Homo sapiens.

PN WO2003003906-A2.

PD 16-JAN-2003.

PF 03-JUL-2002; 2002WO-US21338.

PR 03-JUL-2001; 2001US-302814P.

PR 03-AUG-2001; 2001US-310099P.

PR 08-NOV-2001; 2001US-343705P.

PR 13-NOV-2001; 2001US-350666P.

PR 12-APR-2002; 2002US-372246P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Mack DH, Aziz N;

DR WPI; 2003-201532/19.

DR N-PSDB; ACC51017.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 142 PLDAVECHRGECQAGVLFFQGHGRNCTGHGNSTHHGPYWRCSPHLVLSALTSDNHG 201  
DB 379 KIDAAV-----FNPRFYR----- 391

QY 202 ATYAFSGTHYWRDLTISR-----GHSWPIAHQWPGPSAVDAAFSWEK-LYLVOSTQYV 256  
DB 392 -TYFVNDQYWRDYDERQWMDPGYPLITKNFGIGPK-IDAVFYKKNRYFFQSSNOP 449

QY 257 VF 258  
DB 450 EY 451

## RESULT 13

ABU56421  
ID ABU56421 standard; Protein; 470 AA.

AC ABU56421;

DT 02-APR-2003 (first entry)

DE Lung cancer-associated polypeptide #14.

KW Lung cancer-associated polypeptide; cytostatic; emphysema;

KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;

KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;

KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;

KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

OS Unidentified.

PN WO200286443-A2.

PD 31-OCT-2002.

PF 18-APR-2002; 2002WO-US12476.

PR 18-APR-2001; 2001US-284770P.

PR 10-MAY-2001; 2001US-290492P.

PR 09-NOV-2001; 2001US-339245P.

PR 13-NOV-2001; 2001US-350666P.

PR 29-NOV-2001; 2001US-354370P.

PR 12-APR-2002; 2002US-372246P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Aziz N, Murray R;

DR WPI; 2003-093161/08.

DR N-PSDB; ABX76137.

XX

XX

XX

XX

XX

XX

XX

XX

Claim 27, Page 198; 453p; English.

XX The invention relates to a method for detecting a lung cancer-associated

XX transcript in a cell from a patient, comprising contacting a biological

XX sample from the patient with a polynucleotide that selectively hybridizes

XX to a sequence that is at least 80 % identical to a gene that exhibits

XX increased or decreased expression in lung cancer samples. Lung

XX cancer-associated polynucleotides and polypeptides are used for

XX identifying a compound that modulates a lung cancer-associated

XX polypeptide, for inhibiting proliferation of a lung cancer-associated

XX cell to treat lung cancer in a patient and for treating a mammal having

XX lung cancer by administering a modulatory compound identified. The

XX methods are useful for treating lung cancer, such as small cell lung

XX cancer, non-small cell lung cancer or other benign or precancerous

XX lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive

XX pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial

CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
CC and polypeptides are useful for diagnostic purposes and as targets for  
CC screening for therapeutic compounds that modulate lung cancer, such as  
CC antibodies. Sequences ABUS6408-ABUS6745 represent lung cancer-associated  
CC polypeptides of the invention.

XX Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGTKPPDVT--RCSGWSFDATLLDNGTMLFKGEFVW---KSHKMDR 83  
DB 260 SLVGDPEKNQRLPNDSEPALCPNLSFDAVTTVGN-KIFFKDRFWMKVSERPKTSV 318  
QY 84 ELISERWKNPPSPVDAFR-QGHSNVLKGDKVVVYPEKKEKGYPKLAD-EFGIPS 141  
DB 319 NLISLWPTLPSGIEAAVEIARQVFLPKDKYWLISNLRPEPNYPKSISFGFPNFK 378  
QY 142 PLDAVECHRGECQAEGLFFQGHGHRNGTGHGNSHHGPRVMCSHLVLSALTSNMG 201  
DB 379 KIDAAV-----FNPRPYR----- 391  
QY 202 ATYAFSGTHWRLDTSRD---GWSWPIAHQWPGSPSAVDAAFSWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNDQWRYRDERRQMDPGYPKLITKNFGIGRK-IDAVFYSKKYYTFPGSQNF 449  
QY 257 VF 258  
DB 450 EY 451

RESULT 14

ID ABUS6663 standard; Protein; 470 AA.

AC ABUS6663;

DT 02-APR-2003 (first entry)

DB Lung cancer-associated polypeptide #256.

XX Lung cancer-associated polypeptide; cytostatic; emphysema;  
XX antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

XX Unidentified.

OS WO200286443-A2.

PN 31-OCT-2002.

PD 18-APR-2002; 2002WO-US12476.

PE 18-APR-2001; 2001US-284770P.

PR 10-MAY-2001; 2001US-290492P.

PR 09-NOV-2001; 2001US-339245P.

PR 13-NOV-2001; 2001US-350666P.

PR 29-NOV-2001; 2001US-334370P.

PR 12-APR-2002; 2002US-372246P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

PA Aziz N, Murray R;

PI MPI; 2003-093161/08.

DR N-PBDB; ABX76392.

PT Detecting a lung cancer-associated transcript in a cell from a patient  
for treating lung cancer, by contacting a biological sample from the

PT patient with a polynucleotide that exhibits increased or decreased  
PT expression in lung cancer

PS Claim 27; Page 385-386; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated  
XX transcript in a cell from a patient, comprising contacting a biological  
XX sample from the patient with a polynucleotide that selectively hybridizes  
XX to a sequence that is at least 80 % identical to a gene that exhibits  
XX increased or decreased expression in lung cancer samples. Lung  
XX cancer-associated polynucleotides and polypeptides are used for  
XX identifying a compound that modulates a lung cancer-associated  
XX polypeptide, for inhibiting proliferation of a lung cancer-associated  
XX cell to treat lung cancer in a patient and for treating a mammal having  
XX lung cancer by administering a modulatory compound identified. The  
XX methods are useful for treating lung cancer, such as small cell lung  
XX cancer, non-small cell lung cancer or other benign or precancerous  
XX lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive  
XX pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial  
XX pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
XX and polypeptides are useful for diagnostic purposes and as targets for  
XX screening for therapeutic compounds that modulate lung cancer, such as  
XX antibodies. Sequences ABUS6408-ABUS6745 represent lung cancer-associated  
XX polypeptides of the invention.

Sequence 470 AA;

Query Match 9.3%; Score 202.5; DB 24; Length 470;  
Best Local Similarity 27.3%; Pred. No. 3.9e-10;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAEGTKPPDVT--RCSGWSFDATLLDNGTMLFKGEFVW---KSHKMDR 83  
DB 260 SLVGDPEKNQRLPNDSEPALCPNLSFDAVTTVGN-KIFFKDRFWMKVSERPKTSV 318  
QY 84 ELISERWKNPPSPVDAFR-QGHSNVLKGDKVVVYPEKKEKGYPKLAD-EFGIPS 141  
DB 319 NLISLWPTLPSGIEAAVEIARQVFLPKDKYWLISNLRPEPNYPKSISFGFPNFK 378  
QY 142 PLDAVECHRGECQAEGLFFQGHGHRNGTGHGNSHHGPRVMCSHLVLSALTSNMG 201  
DB 379 KIDAAV-----FNPRPYR----- 391  
QY 202 ATYAFSGTHWRLDTSRD---GWSWPIAHQWPGSPSAVDAAFSWEK-LYLVQGTQVY 256  
DB 392 -TYFFVNDQWRYRDERRQMDPGYPKLITKNFGIGRK-IDAVFYSKKYYTFPGSQNF 449  
QY 257 VF 258  
DB 450 EY 451

RESULT 15

ID AAB43772 standard; Protein; 473 AA.

AC AAB43772;

DT 08-FEB-2001 (first entry)

DB Human cancer associated protein sequence SEQ ID NO:1217.

XX Human, cancer associated gene; cancer antigen; detection; cancer;  
XX diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;  
XX antidiabetic; antiasthmatic; antirheumatic; antibrutic; antiviral;  
XX antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
XX dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
XX vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
XX immune disorder; haematopoietic cell disorder; autoimmune disorder;  
XX allergic reaction; graft versus host disease; organ rejection;  
XX haemostatic; thrombolytic; cardiovascular disorder; infection;  
XX neurological disease; drug screening.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 16, 2003, 06:46:44 ; Search time 28 Seconds  
(without alignments)  
590.841 Million cell updates/sec

Title: US-09-900-448-2  
Perfect score: 2185  
Sequence: 1 MARVLGAPVALGLWSLWSL.....NAKALPQPVNTSLGCTH 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	202.5	9.3	470	US-08-068-392-2	Sequence 2, Appl1
2	202.5	9.3	470	US-08-396-988-2	Sequence 2, Appl1
3	202.5	9.3	470	US-09-391-104-26	Sequence 26, Appl1
4	199	9.1	517	US-09-391-104-32	Sequence 32, Appl1
5	199	9.1	519	US-09-211-704A-7	Sequence 7, Appl1
6	189.5	8.7	476	US-08-704-711A-21	Sequence 21, Appl1
7	189.5	8.7	476	US-08-448-489-14	Sequence 14, Appl1
8	189.5	8.7	476	US-09-521-220-21	Sequence 21, Appl1
9	189.5	8.7	476	US-09-391-104-22	Sequence 22, Appl1
10	183.5	8.4	467	US-09-178-002-2	Sequence 2, Appl1
11	183.5	8.4	467	US-09-391-104-24	Sequence 4, Appl1
12	183.5	8.4	467	US-08-448-489-13	Sequence 13, Appl1
13	183.5	8.4	468	US-07-794-393-4	Sequence 23, Appl1
14	181.5	8.3	492	US-07-794-393-4	Sequence 4, Appl1
15	181.5	8.3	492	US-08-001-711-4	Sequence 2, Appl1
16	181.5	8.3	488	US-07-794-393-2	Sequence 2, Appl1
17	179.5	8.2	488	US-08-001-711-2	Sequence 2, Appl1
18	179.5	8.2	488	US-08-704-711A-22	Sequence 22, Appl1
19	179.5	8.2	488	US-09-391-104-31	Sequence 31, Appl1
20	179.5	8.2	488	US-08-448-489-11	Sequence 11, Appl1
21	179.5	8.2	489	US-08-448-489-15	Sequence 15, Appl1
22	179.5	8.2	477	US-08-704-711A-20	Sequence 20, Appl1
23	177	8.1	477	US-08-448-489-15	Sequence 15, Appl1
24	177	8.1	477	US-08-281-313-1	Sequence 9, Appl1
25	177	8.1	477	US-09-521-220-20	Sequence 20, Appl1
26	177	8.1	477	US-09-391-104-21	Sequence 21, Appl1
27	177	8.1	477		

28	176.5	8.1	469	US-08-704-711A-16	Sequence 16, Appl1
29	176.5	8.1	469	US-08-448-489-12	Sequence 12, Appl1
30	176.5	8.1	469	US-09-521-220-16	Sequence 16, Appl1
31	176.5	8.1	471	US-09-391-104-25	Sequence 25, Appl1
32	175	8.0	466	US-08-704-711A-17	Sequence 17, Appl1
33	175	8.0	466	US-09-521-220-17	Sequence 17, Appl1
34	172	7.9	564	US-09-211-704A-8	Sequence 8, Appl1
35	172	7.9	669	US-08-704-711A-3	Sequence 3, Appl1
36	172	7.9	669	US-09-521-220-3	Sequence 3, Appl1
37	172	7.9	669	US-09-391-104-29	Sequence 29, Appl1
38	169.5	7.8	508	US-09-391-104-18	Sequence 18, Appl1
39	167.5	7.7	411	US-09-171-545-3	Sequence 3, Appl1
40	167.5	7.7	416	US-09-171-545-4	Sequence 4, Appl1
41	167.5	7.7	508	US-09-171-545-1	Sequence 1, Appl1
42	164.5	7.5	324	US-08-816-755-2	Sequence 2, Appl1
43	164.5	7.5	324	US-09-090-673-2	Sequence 2, Appl1
44	161.5	7.4	532	US-09-294-841-2	Sequence 2, Appl1
45	154	7.0	419	US-09-171-545-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1  
US-08-068-392-2  
; Sequence 2, Application US/08068392  
; Patent No. 6150152  
; GENERAL INFORMATION:  
; APPLICANT: Shapito, Steven M.  
; TITLE OF INVENTION: Human Macrophage Metalloproteinase  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A35M  
; STREET: 800 N. Lindbergh Blvd.  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/068,392  
; FILING DATE: 19930528  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyer, Scott J.  
; REGISTRATION NUMBER: 25275  
; REFERENCE/DOCKET NUMBER: 07-24(12406)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3117  
; TELEFAX: (314)694-5435  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 470 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-068-392-2

Query Match 9.3%; Score 202.5; DB 3; Length 470;  
Best Local Similarity 27.3%; Pred. No. 2.4e-12;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;

QY 30 SAHGNVAAGETKPPDYTE--RCSDGSPDATTDDNGTMLFFKGFVW----KSHKWRD 83  
DB 260 SYXDPPKRNQRLPRPDSEPLCDPNLSFDATVTGN-KIFFPDRFMLKVSRRPTSV 318  
QY 84 ELISERKMPSPVDAAR-QGHSVFLIKDKRWVVPPEKKEGYPRLQD-EFGIGPS 141  
DB 319 NLISLWPTLPSGIEAAVIEARNOVFLFKDKKYLISNLRPBNYPKSIHSFGFPNFKV 378



QY 142 PLDAVECHRGCOAGVLFPOGHGHRNGTGHGNSTHHGPEYMRCSPHVLVLSALTSDNHG 201  
| : : : : :  
DB 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD---GWHSWPIAHQWPQGPSAVDAAFSWEER-LYLVQGTQVY 256  
| : : : : :  
DB 392 -TYFFVNDQYWRDYDERQWMDPGYPKLITKNQIGPK-IDAVFYSKNKYFFFGQSNQF 449  
QY 257 VF 258  
DB 450 EY 451  
RESULT 2  
US-08-396-988-2  
Sequence 2, Application US/08396988  
Patent No. 6204043  
GENERAL INFORMATION:  
APPLICANT: Shapiro, Steven M.  
TITLE OF INVENTION: Human Macrophage Metalloproteinase  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SM  
STREET: 800 N. Lindbergh Blvd.  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63167  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,988  
FILING DATE: 01-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/068,392  
FILING DATE: 28-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyer, Scott J.  
REGISTRATION NUMBER: 25275  
REFERENCE/DOCKET NUMBER: 07-24(12406)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)694-3117  
TELEFAX: (314)694-5435  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-988-2  
Query Match 9.3%; Score 202.5; DB 3; Length 470;  
Best Local Similarity 27.3%; Pred. No. 2.4e-12;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
QY 30 SAHGNAVEGETKPDPPVTE--RCSDSGSPDATLLDNGTMLPFKGEFVW----KSHKMDR 83  
| : : : : :  
DB 260 SLVGDPKRNQRLNPNPNSPALCDPNLSFDAVTVVGN-KIFFKDRFFMLKXSERKTSV 318  
QY 84 ELISERKNFPPSPVDAFR-QGHNSVFLIKGKRWVYPPKKEKGYPKLIQD-EFGIGPS 141  
| : : : : :  
DB 319 NLISSLMPTLPSPGIEAAYEIEARNQVFLFKDXYKWLISNLRBPENPKSIHSFGFNFVK 378  
QY 142 PLDAVECHRGCOAGVLFPOGHGHRNGTGHGNSTHHGPEYMRCSPHVLVLSALTSDNHG 201  
| : : : : :  
DB 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD---GWHSWPIAHQWPQGPSAVDAAFSWEER-LYLVQGTQVY 256  
| : : : : :  
DB 450 EY 451

DB 392 -TYFFVNDQYWRDYDERQWMDPGYPKLITKNQIGPK-IDAVFYSKNKYFFFGQSNQF 449  
QY 257 VF 258  
DB 450 EY 451  
RESULT 3  
US-09-391-104-26  
Sequence 26, Application US/09391104  
Patent No. 6399371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Falduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
TITLE OF INVENTION: OF USING SAME  
FILE REFERENCE: 6073.US.P1  
CURRENT APPLICATION NUMBER: US/09/391,104  
CURRENT FILING DATE: 1999-09-07  
PRIOR APPLICATION NUMBER: US 08/814,394  
PRIOR FILING DATE: 1997-03-11  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 26  
LENGTH: 470  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-391-104-26  
Query Match 9.3%; Score 202.5; DB 4; Length 470;  
Best Local Similarity 27.3%; Pred. No. 2.4e-12;  
Matches 66; Conservative 23; Mismatches 90; Indels 63; Gaps 10;  
QY 30 SAHGNAVEGETKPDPPVTE--RCSDSGSPDATLLDNGTMLPFKGEFVW----KSHKMDR 83  
| : : : : :  
DB 260 SLVGDPKRNQRLNPNPNSPALCDPNLSFDAVTVVGN-KIFFKDRFFMLKXSERKTSV 318  
QY 84 ELISERKNFPPSPVDAFR-QGHNSVFLIKGKRWVYPPKKEKGYPKLIQD-EFGIGPS 141  
| : : : : :  
DB 319 NLISSLMPTLPSPGIEAAYEIEARNQVFLFKDXYKWLISNLRBPENPKSIHSFGFNFVK 378  
QY 142 PLDAVECHRGCOAGVLFPOGHGHRNGTGHGNSTHHGPEYMRCSPHVLVLSALTSDNHG 201  
| : : : : :  
DB 379 KIDAAV-----FNRFYR----- 391  
QY 202 ATYAFSGTHYWRDLTSDRD---GWHSWPIAHQWPQGPSAVDAAFSWEER-LYLVQGTQVY 256  
| : : : : :  
DB 392 -TYFFVNDQYWRDYDERQWMDPGYPKLITKNQIGPK-IDAVFYSKNKYFFFGQSNQF 449  
QY 257 VF 258  
DB 450 EY 451  
RESULT 4  
US-09-391-104-32  
Sequence 32, Application US/09391104  
Patent No. 6399371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Falduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
TITLE OF INVENTION: OF USING SAME  
FILE REFERENCE: 6073.US.P1  
CURRENT APPLICATION NUMBER: US/09/391,104  
CURRENT FILING DATE: 1999-09-07

PRIOR APPLICATION NUMBER: US 08/814,394  
 PRIOR FILING DATE: 1997-03-11  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 32  
 LENGTH: 517  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-391-104-32

Query Match 9.1%; Score 199; DB 4; Length 517;  
 Best local similarity 25.0%; Pred. No. 6.3e-12;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----WSLCWSLAIATPL-----PTSAHGNVAEGETKPDPTVTERCS 51  
 DB 188 VGDPLRYGLPEYDKRVWQLYGVRESVPTAQPEEPPLPEPPDNRSAPPRKDVPHRCS 247  
 QY 52 DGWSPDATLDDNGTMTLFFKGEFVWKSXKMDRELISER-----WKNFP---SPVDAAF 101  
 DB 248 T--HFDAVA-QIRGEAFFFKGYFWRLTR-DRHLVSLQPAQMHRFWRLPLHLDSVDAYV 303  
 QY 102 -RQGHNSVFLIKGDKVWVYPPPEKKEKGYPKLLODFPGIPEPLDAAVECHRGECQAGSVL 160  
 DB 304 ERTSDHKIVFFKGRVYVFKDNVVEEGYPRVSD-FSLPPGIDAA----- 348  
 QY 161 FFQGHGRNGTGHGNSTHGPEVWRCSPHLVLSALTSNMGATYAFSGTHYWRDLT--SRD 219  
 DB 349 FSNMHNDR-----TYFFKDQLYWRVDHTRH 374  
 QY 220 GMSWPI-AHQWPGPSAVDAAFSWEE-KLYLVQGTQVYVFLTKGYTLVSGYPRLEKE 277  
 DB 375 MDGYPAGSPLMRGVPSLDDAMRWSDGASYFFRG-QEYWKVLDGELEVAAGYQSTARD 433  
 QY 278 -----VGTPHGIILDSVDA 292  
 DB 434 WLVCQDSQADGSVAAGVDA 453

RESULT 5

US-09-211-704A-7  
 Sequence 7, Application US/09211704A  
 Patent No. 6271014  
 GENERAL INFORMATION:  
 APPLICANT: de Saint-Vis, Blandine Marie  
 APPLICANT: Fossiez, Francois  
 APPLICANT: Caux, Christophe  
 APPLICANT: Lebecque, Serge J.E.  
 TITLE OF INVENTION: Mammalian Proteinases; Related Reagents  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DNAX Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/211,704A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/005,263  
 FILING DATE: 09-JAN-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,090

REFERENCE/DOCKET NUMBER: SF0781K  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 852-9196  
 TELEFAX: (650) 496-1200  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 519 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-211-704A-7

Query Match 9.1%; Score 199; DB 3; Length 519;  
 Best local similarity 25.0%; Pred. No. 6.4e-12;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

QY 5 LGAPVALGL-----WSLCWSLAIATPL-----PTSAHGNVAEGETKPDPTVTERCS 51  
 DB 190 VGDPLRYGLPEYDKRVWQLYGVRESVPTAQPEEPPLPEPPDNRSAPPRKDVPHRCS 249  
 QY 52 DGWSPDATLDDNGTMTLFFKGEFVWKSXKMDRELISER-----WKNFP---SPVDAAF 101  
 DB 250 T--HFDAVA-QIRGEAFFFKGYFWRLTR-DRHLVSLQPAQMHRFWRLPLHLDSVDAYV 305  
 QY 102 -RQGHNSVFLIKGDKVWVYPPPEKKEKGYPKLLODFPGIPEPLDAAVECHRGECQAGSVL 160  
 DB 306 ERTSDHKIVFFKGRVYVFKDNVVEEGYPRVSD-FSLPPGIDAA----- 350  
 QY 161 FFQGHGRNGTGHGNSTHGPEVWRCSPHLVLSALTSNMGATYAFSGTHYWRDLT--SRD 219  
 DB 351 FSNMHNDR-----TYFFKDQLYWRVDHTRH 376  
 QY 220 GMSWPI-AHQWPGPSAVDAAFSWEE-KLYLVQGTQVYVFLTKGYTLVSGYPRLEKE 277  
 DB 377 MDGYPAGSPLMRGVPSLDDAMRWSDGASYFFRG-QEYWKVLDGELEVAAGYQSTARD 435  
 QY 278 -----VGTPHGIILDSVDA 292  
 DB 436 WLVCQDSQADGSVAAGVDA 455

RESULT 6

US-08-704-711A-21  
 Sequence 21, Application US/08704711A  
 Patent No. 6114159  
 GENERAL INFORMATION:  
 APPLICANT: WILZ, Horst  
 APPLICANT: HINZMANN, Bernd  
 TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/704,711A  
 FILING DATE: 20-NOV-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/DE95/00357  
 FILING DATE: 17-MAR-1995  
 APPLICATION NUMBER: DE 443838.1

FILING DATE: 21-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-704-711A-21

Query Match 8.7%; Score 189.5; DB 3; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

Qy 25 PLPPTSAHGVNABGETKPDVTERCSDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKMDR 83  
Db 272 PLVPTK---SVPGSEMP-----AKCDPALSPDAISTL--RGEYLFKDRYFWRSHMNP 321

Qy 84 E-----LISRWKNPSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120  
Db 322 EPEFHLSAFWPSLPSYLDAAVEVNSRDYVIFKGNFPAIRGNEVOAGYPRGIHTLGFP 381

Qy 121 P-----EKK-----EKGYPKLQDEFGIPSPDLAAVE 148  
Db 382 PTIRKIDAAVSDKEKKTYFFAADKTYRFDENSQSMEOGFPRLIADDFGVEPKVDVAVL- 440

Qy 149 CHRGECCAGVLEFP 162  
Db 441 -----QAFGFFYF 448

RESULT 7  
US-08-448-489-14  
Sequence 14, Application US/08448489  
Patent No. 6184022  
GENERAL INFORMATION:  
APPLICANT: SEIKI, Motoharu  
APPLICANT: SATO, Hiroshi  
APPLICANT: SHINAGAWA, Akira  
TITLE OF INVENTION: NOVEL METALLOPROTEINASE AND ENCODING DNA THEREFOR  
FILE REFERENCE: 55-280P  
CURRENT APPLICATION NUMBER: US/08/448,489  
CURRENT FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Known Member of  
OTHER INFORMATION: Matrix Metalloproteinase Family  
US-08-448-489-14

Query Match 8.7%; Score 189.5; DB 3; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

Qy 25 PLPPTSAHGVNABGETKPDVTERCSDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKMDR 83  
Db 272 PLVPTK---SVPGSEMP-----AKCDPALSPDAISTL--RGEYLFKDRYFWRSHMNP 321  
Qy 84 E-----LISRWKNPSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120

Db 322 EPEFHLSAFWPSLPSYLDAAVEVNSRDYVIFKGNFPAIRGNEVOAGYPRGIHTLGFP 381  
Qy 121 P-----EKK-----EKGYPKLQDEFGIPSPDLAAVE 148  
Db 382 PTIRKIDAAVSDKEKKTYFFAADKTYRFDENSQSMEOGFPRLIADDFGVEPKVDVAVL- 440  
Qy 149 CHRGECCAGVLEFP 162  
Db 441 -----QAFGFFYF 448

RESULT 8  
US-09-521-220-21  
Sequence 21, Application US/09521220  
Patent No. 639348  
GENERAL INFORMATION:  
APPLICANT: WILLY, Horst  
HINZMANN, Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX  
METALLOPROTEASES, THEIR PRODUCTION AND USE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESS: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/521,220  
FILING DATE: 08-Mar-2000  
CLASSIFICATION: <Unknown>  
21-OCT-1994  
17-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/704,711  
FILING DATE: <Unknown>  
APPLICATION NUMBER: DE 4438838.1  
FILING DATE: 21-OCT-1994  
APPLICATION NUMBER: DE 4409663.1  
FILING DATE: 17-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 26083/124  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 21:  
US-09-521-220-21

Query Match 8.7%; Score 189.5; DB 4; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

Qy 25 PLPPTSAHGVNABGETKPDVTERCSDGWSFDA-TTLDNGTMTLFFKGEFVWKSHKMDR 83  
Db 272 PLVPTK---SVPGSEMP-----AKCDPALSPDAISTL--RGEYLFKDRYFWRSHMNP 321  
Qy 84 E-----LISRWKNPSPVDAAFR-QGHNVSFLIKGDKVWV-----YP 120

Db 322 EPEFLISAFWPSLPSYLDAAEVNSRDTVFIFKGNEMFMAIRGNEVQAGYPRGITHLGF 381  
Qy 121 P-----EKK-----EKGYPLLODEFPFGIPSPDLDAVE 148  
Db 382 PTIRKIDAAVSDKEKKKTYFFPAADKYRWPFDENSQSGFPPLIADDFGVPEPKVDALV 440  
Qy 149 CHRGECAQEGVLFF 162  
Db 441 -----QAFGFFVF 448

RESULT 9  
US-09-391-104-22

; Sequence 22, Application US/09391104  
; Patent No. 6399371  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Falduto, Michael T.  
; APPLICANT: Magnuson, Scott R.  
; APPLICANT: Morgan, Douglas W.  
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
; TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
; FILE REFERENCE: 6073.US.P1  
; CURRENT APPLICATION NUMBER: US/09/391,104  
; CURRENT FILING DATE: 1999-09-07  
; PRIOR APPLICATION NUMBER: US 08/814,394  
; PRIOR FILING DATE: 1997-03-11  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 22  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-391-104-22

Query Match 8.7%; Score 189.5; DB 4; Length 476;  
Best Local Similarity 28.4%; Pred. No. 5.4e-11;  
Matches 55; Conservative 24; Mismatches 42; Indels 73; Gaps 10;

Qy 25 PLPPTSAHGNVAEGTKPDPTVTERCSDGWSFDA-TTLDNGTMLFFKGEFVWKS HKWDR 83  
Db 272 PLVPTK---SVPSGSEMP-----AKCDPALSFDAISTL--RGEYLFKDRYFWRSHNRP 321  
Qy 84 E---LISERKNPSPVDAAFR--OGHNSVFLIKGDKVW-----YP 120  
Db 322 EPEFLISAFWPSLPSYLDAAEVNSRDTVFIFKGNEMFMAIRGNEVQAGYPRGITHLGF 381  
Qy 121 P-----EKK-----EKGYPLLODEFPFGIPSPDLDAVE 148  
Db 382 PTIRKIDAAVSDKEKKKTYFFPAADKYRWPFDENSQSGFPPLIADDFGVPEPKVDALV 440  
Qy 149 CHRGECAQEGVLFF 162  
Db 441 -----QAFGFFVF 448

RESULT 10  
US-09-178-002-2

; Sequence 2, Application US/09178002  
; Patent No. H001973  
; GENERAL INFORMATION:  
; APPLICANT: Hu, Shou-ih  
; TITLE OF INVENTION: Human Neutrophil Collagenase Splice Variant  
; FILE REFERENCE: CGC 2048  
; CURRENT APPLICATION NUMBER: US/09/178,002  
; CURRENT FILING DATE: 1998-10-22  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 444  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-09-178-002-2

Query Match 8.4%; Score 183.5; DB 1; Length 444;  
Best Local Similarity 24.4%; Pred. No. 2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

Qy 23 ATPPPTSAHGNVAEGTKPDPTVTERCSDGWSFDA-TTLDNGTMLFFKGEFVWKS H-- 79  
Db 242 SNPIQPT-----GSTPKP-----CDPSLTFDAITTL--RGEILFFKDRYFWRSHNPQ 286  
Qy 80 --KMDRELISERKNPSPVDAAFRQ--GHSVFLIKGDKVWVPEKKEKGYPLLOD-E 135  
Db 287 LQREVMNFIPLFWPSLPTGIGAAVEDFRDLIFLFGKNGQWALSQYDILQGYPRDISNYG 346  
Qy 136 FPGIPSPDLDAVECHRGECAQEGVLFFQGHGHRNGTGHGNSHTHGPEYMRCS PHLVLSAL 195  
Db 347 FPSSVQAIDAAV-----FYR----- 361  
Qy 196 TSDNHGATVAFSGTHYWRDLTSR---DGMHSWPIAHQWPGPSAVDAAFSMEKLYLVQG 252  
Db 362 ----SKTYFVNDQFWRYNQRFMEPGYPKSISGAPGIESKVDVAFQOEHFHFVPSG 416  
Qy 253 TOYTVF 258  
Db 417 PRYVAF 422

RESULT 11  
US-09-178-002-4

; Sequence 4, Application US/09178002  
; Patent No. H001973  
; GENERAL INFORMATION:  
; APPLICANT: Hu, Shou-ih  
; TITLE OF INVENTION: Human Neutrophil Collagenase Splice Variant  
; FILE REFERENCE: CGC 2048  
; CURRENT APPLICATION NUMBER: US/09/178,002  
; CURRENT FILING DATE: 1998-10-22  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 467  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-178-002-4

Query Match 8.4%; Score 183.5; DB 1; Length 467;  
Best Local Similarity 24.4%; Pred. No. 2.2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

Qy 23 ATPPPTSAHGNVAEGTKPDPTVTERCSDGWSFDA-TTLDNGTMLFFKGEFVWKS H-- 79  
Db 265 SNPIQPT-----GSTPKP-----CDPSLTFDAITTL--RGEILFFKDRYFWRSHNPQ 309  
Qy 80 --KMDRELISERKNPSPVDAAFRQ--GHSVFLIKGDKVWVPEKKEKGYPLLOD-E 135  
Db 310 LQREVMNFIPLFWPSLPTGIGAAVEDFRDLIFLFGKNGQWALSQYDILQGYPRDISNYG 369  
Qy 136 FPGIPSPDLDAVECHRGECAQEGVLFFQGHGHRNGTGHGNSHTHGPEYMRCS PHLVLSAL 195  
Db 370 FPSSVQAIDAAV-----FYR----- 384  
Qy 196 TSDNHGATVAFSGTHYWRDLTSR---DGMHSWPIAHQWPGPSAVDAAFSMEKLYLVQG 252  
Db 385 ----SKTYFVNDQFWRYNQRFMEPGYPKSISGAPGIESKVDVAFQOEHFHFVPSG 439  
Qy 253 TOYTVF 258  
Db 440 PRYVAF 445

RESULT 12  
US-09-391-104-24

Sequence 24, Application US/09391104  
Patent No. 6399371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Falduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
TITLE OF INVENTION: OF USING SAME  
FILE REFERENCE: 6073.US.P1  
CURRENT APPLICATION NUMBER: US/09/391.104  
CURRENT FILING DATE: 1999-09-07  
PRIOR APPLICATION NUMBER: US 08/814,394  
PRIOR FILING DATE: 1997-03-11  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FaastSeq for Windows Version 3.0  
SEQ ID NO 24  
LENGTH: 467  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-391-104-24

Query Match 8.4%; Score 183.5; DB 4; Length 467;  
Best Local Similarity 24.4%; Pred. No. 2.2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

QY 23 ATPLPPTSAHGNVAEGTETDPDPTERTCSGDSFDA-TTLDNGTMLFFKGEFVWKSH-- 79  
DB 265 SNPIQPT-----GSTPKP-----CDPSLTFDAITTL--RGEILFFDRYFWRHHPQ 309  
QY 80 --KMDRELISERKWNKPPSYDAAFRO-GHNSVFLIKGDKVWVYPRKKEGYPKLQD-E 135  
DB 310 LQEVNMFISLFWPSLPTGTGIAAYEDPDRLIFLFGKNQWALSGYDILQGYPKDISNG 369  
QY 136 FPGIPSPDLAAVECHRGCEQAEGLFFQGHGHRNGTGHNSTHGPEYRCSPLVLAL 195  
DB 370 FPGSPQALDAV-----FYR----- 384  
QY 196 TSDNHGATTAFASTGTHWRDLTSR---DGMHSPRIAHQWPGPSAVDAAFSWEKLYLVQG 252  
DB 385 -----SKTYFFVNDQFWRVNDQRFMEPGYPKSISGAFPGIESKVDVAFQOEHPFHVFG 439  
QY 253 TOYVVF 258  
DB 440 PRYYAF 445

## RESULT 13

US-08-448-489-13  
Sequence 13, Application US/08448489  
Patent No. 6184022  
GENERAL INFORMATION:  
APPLICANT: SEIKI, Motoharu  
APPLICANT: SATO, Hiroshi  
APPLICANT: SHINAGAWA, Akira  
TITLE OF INVENTION: NOVEL METALLOPROTEINASE AND ENCODING DNA THEREFOR  
FILE REFERENCE: 55-290P  
CURRENT APPLICATION NUMBER: US/08/448,489  
CURRENT FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 468  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: X = UNKNOWN  
OTHER INFORMATION: Description of Unknown Organism: Known Member of  
OTHER INFORMATION: Matrix Metalloproteinase Family  
US-08-448-489-13

Query Match 8.4%; Score 183.5; DB 3; Length 468;  
Best Local Similarity 24.4%; Pred. No. 2.2e-10;  
Matches 60; Conservative 31; Mismatches 80; Indels 75; Gaps 10;

QY 23 ATPLPPTSAHGNVAEGTETDPDPTERTCSGDSFDA-TTLDNGTMLFFKGEFVWKSH-- 79  
DB 265 SNPIQPT-----GSTPKP-----CDPSLTFDAITTL--RGEILFFDRYFWRHHPQ 309  
QY 80 --KMDRELISERKWNKPPSYDAAFRO-GHNSVFLIKGDKVWVYPRKKEGYPKLQD-E 135  
DB 310 LQEVNMFISLFWPSLPTGTGIAAYEDPDRLIFLFGKNQWALSGYDILQGYPKDISNG 369  
QY 136 FPGIPSPDLAAVECHRGCEQAEGLFFQGHGHRNGTGHNSTHGPEYRCSPLVLAL 195  
DB 370 FPGSPQALDAV-----FYR----- 384  
QY 196 TSDNHGATTAFASTGTHWRDLTSR---DGMHSPRIAHQWPGPSAVDAAFSWEKLYLVQG 252  
DB 385 -----SKTYFFVNDQFWRVNDQRFMEPGYPKSISGAFPGIESKVDVAFQOEHPFHVFG 439  
QY 253 TOYVVF 258  
DB 440 PRYYAF 445

## RESULT 14

US-09-391-104-23  
Sequence 23, Application US/09391104  
Patent No. 6399371  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
APPLICANT: Falduto, Michael T.  
APPLICANT: Magnuson, Scott R.  
APPLICANT: Morgan, Douglas W.  
TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,  
TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS  
TITLE OF INVENTION: OF USING SAME  
FILE REFERENCE: 6073.US.P1  
CURRENT APPLICATION NUMBER: US/09/391.104  
CURRENT FILING DATE: 1999-09-07  
PRIOR APPLICATION NUMBER: US 08/814,394  
PRIOR FILING DATE: 1997-03-11  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FaastSeq for Windows Version 3.0  
SEQ ID NO 23  
LENGTH: 469  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-391-104-23

Query Match 8.3%; Score 181.5; DB 4; Length 469;  
Best Local Similarity 26.3%; Pred. No. 3.6e-10;  
Matches 59; Conservative 26; Mismatches 82; Indels 57; Gaps 8;

QY 44 PDVTERCSGDSFDA-TTLDNGTMLFFKGEFVWKSHKMDREL-----ISERKWNKPPSYD 98  
DB 272 PQPKACDSGLTFDAITTL--RGEVWFPRDRFYMRNPPRYVELNLFISVFWQLNGSE 329  
QY 99 AAFR-OGHNSVFLIKGDKVWVYPRKKEGYPKLQDDEPPGIPSPDLAAVECHRGCEQAE 157  
DB 330 AAYEFADRDVEVRFKKNKYMAVOGQVVLHGYPKDIYSF-GFP----- 371  
QY 158 GVLFFQGHGHRNGTGHNSTHGPEYRCSPLVLALSDNHGATTAFASTGTHWRDLTSR 215  
DB 372 -----RTVKH-IDALSEENGTFFVANKYRHYDEY 403  
QY 216 -TSRDGMSWPIAHQWPGPSAVDAAFSWEKLYLVQGTQVVF 258  
DB 404 KRSMDPGYRMLAHDPFGIGHKVDVAFMDGPFYFFHGTROYVF 447

## RESULT 15

US-07-794-393-4

Sequence 4, Application US/0794393  
Patent No. 5236844  
GENERAL INFORMATION:  
APPLICANT: CHAMON, PIERRE  
APPLICANT: BASSET, PAUL  
APPLICANT: BELLOCQ, JEAN-PIERRE  
TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST  
TITLE OF INVENTION: CANCER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
Street: 1225 Connecticut Ave. NW Suite 300  
City: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/07/794,393  
APPLICATION NUMBER: US/07/794,393  
FILING DATE: 19911121  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9025326.1  
FILING DATE: 21-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: GOLDSTEIN, JORGE A  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 1383.0040000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-794-393-4

Query Match 8.3%; Score 181.5; DB 1; Length 492;  
Best Local Similarity 25.2%; Pred. No. 3.8e-10;  
Matches 69; Conservative 26; Mismatches 86; Indels 93; Gaps 13;

QY 24 TPLPPTSAH-----GNVAEGSTKPDPDYTERCSGMSFDPATLLDNGTMLFFKGFVWKS 78  
Db 270 SPAPTLSSQAGDTNEMIALLEPETPPDVCET-----SFDAYS-TIRGELFFKAGFWRL 323  
QY 79 HKMDRE-----LISRWKGNFSPVDAAFRQGHNSVFLIKDKVWYYPPEKKEKGYPKLLQ 133  
Db 324 RSGRLQPGYPALASHRWGLPSVDAAFEADAQGIWFFQAGYWWYDDEKPLGPAPLSK 383  
QY 134 DEFGIPSPPLDAAVECHRGEGQAEGVLFQGHGHRNGTHGNSTHGPEYWRCSPHVLVS 193  
Db 384 LGLOG--SPVAALV-----WGPE-----400  
QY 194 ALTSDNHGATYAFSGTHYRLD--TSR-DGWHSWPIAHQ--WPGPSAVDAAFSWEKL 247  
Db 401 -----KKIYFFPGGDMRPHRPTQAVDN---PVPRRSTDMRGVPSSEIDAF-----444  
QY 248 YLVGCTQYVFLTKG-----GYTVSGYPK 272  
Db 445 ---QDAEGYAYFLRGLHYMKFDPVKVYKLGSPR 475

Search completed: December 16, 2003, 06:55:24  
Job time : 29 secs

**THIS PAGE BLANK (uspto)**





```

DB 121 PEKKEKGYKLLQDEFFPGIPSPUDAAVECHRGCEQAGVLFPGQDREMFMDLATGTWKEK 180
QY 164 -----GCHGRN 169
DB 181 SWPAVNCSSALRMWLRGYCGQGNQLRFDPRVGEVPPRYPRDVRDYFMPGCRGHGRN 240
QY 170 GTGHGNS 176
DB 241 GTGHGNS 247

RESULT 2
08KJ16 PRELIMINARY; PRT; 276 AA.
ID 08KJ16
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Hemopexin (Fragment)
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RA Alam J., Smith A.;
RT "Nucleotide sequence of the mouse haemopexin gene."
RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; X56829; CAA40160.1;
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF00045; hemopexin; 3.
DR SMART; SM00120; HX; 3.
FT NON TER 276
SQ SEQUENCE 276 AA; 31047 MW; BD336C649BB2B767 CRC64;

Query Match 31.6%; Score 689.5; DB 11; Length 276;
Best Local Similarity 53.0%; Pred. No. 7.8e-50;
Matches 149; Conservative 12; Mismatches 41; Indels 79; Gaps 6;

QY 1 MAWVGLAPVALGMSLCMSLAITPLPPTSAHGNVAGE--TRPDVYTRGSDGNSFDA 58
DB 1 MARTAVANLIVLGLCWSLAVASPLP--TAHGRVAEVENGTXPDSVPBHCIDTWSFDA 58
QY 59 TTLDGDTMLFFGGEFFWVKSHKMDRELISERWKNPSPVDAAFRQGNSTFELIKGDKVWY 118
DB 59 ATDHDHGTMLFFGGEFFWVKSHKMDRELISERWKNPSPVDAAFRQGNSTFELIKGDKVWY 117
QY 119 YPEKKEKGYKLLQDEFFPGIPSPUDAAVECHRGCEQAGVLFPGQDREMFMDLATGTWKEK 165
DB 118 YPEKKEKGYKLLQDEFFPGIPSPUDAAVECHRGCEQAGVLFPGQDREMFMDLATGTWKEK 177
QY 166 -----GCHGRN 169
DB 178 ERSWPAVNCSSALRMWLRGYCGQGNQLRFDPRVGEVPPRYPRDVRDYFMPGCRGHGRN 237
QY 168 -RNGTGHGNSHTHGPXEMRCSPLVLSALTSNDHGATYAPS 207
DB 238 PRNGTGHGNSHTHGPXEMRCSPLVLSALTSNDHGATYAPS 276

RESULT 3
08J1P8 PRELIMINARY; PRT; 427 AA.
ID 08J1P8
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Warm-temperature-acclimation-related-65 kDa-protein-like protein.
GN WAP65-LIKE
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;

```

```

OC Belontiiformes; Adrianichthyidae; Oryziatidae; Oryzias.
RN NCBI_TaxID=8090;
RP SEQUENCE FROM N.A.
RA KINOSHITA S., KOBAYAMA A., KINOSHITA S., WATABE S.;
RT "Primary structural analysis of medaka Wap65 and Wap65-like protein
RT and their gene expression in association with temperature acclimation
RT and during ontogeny."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB075199; BAB97304.1;
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF00045; hemopexin; 3.
DR SMART; SM00120; HX; 7.
SQ SEQUENCE 427 AA; 48000 MW; DEDB66A5BAD38EB CRC64;

Query Match 28.6%; Score 625; DB 13; Length 427;
Best Local Similarity 33.0%; Pred. No. 3.5e-44;
Matches 146; Conservative 66; Mismatches 136; Indels 94; Gaps 12;

QY 15 SLCSLSAATPLPPTSAHGNVAGEGK--PDVYTRGSDGNSPDATLTDNGTMLFFKGE 73
DB 11 ALMLALTTRAAPLEDSAA---GDGSAALPDRCGA-----GIEFDATTPDDKQTFEFGKD 60
QY 74 FVWKSMDRELISERWKNPSPVDAAFR-----QG-HNSVFLIKGRVWVYPPKKEK 126
DB 61 HWKGFGEAGQSSQYFKLNGHVDAFRMHNPNENGDDHDIYLPDDKVFSEFHTLBE 120
QY 127 GYPKLLQDEFFPGIPSPUDAAVECHRGCEQAGVLFPGQDREMFMDLATGTWKEK 164
DB 121 GYPKLEIQEDFPVPAHLDAAVECPKGECTDVLFEKGDVYVYDVTYKTKTWPHLP 180
QY 165 -----HGR-----NGTGHGNSHTHGPXEMRCSPLVLSALTSNDHGATYAPS 240
DB 181 ACTSVFRLWELHYCCFHHNPTFPQVYTGVTGYPKDAARRFYRRCDFGHGERTTLKCS 240
QY 189 HVLVSLATSDNMGATYAFSGTYWRLDTSRDGWSHSPRIAHQPOGSAVDAAFSEWEEKY 248
DB 241 DFMMDAITTDDGGRMYWFMFGSNVRLDTRDGHARPIITSMKELNGDAVAFSYNDRIY 300
QY 249 LVQGTQVYVFLTKGTYLVSIGYPRKLEKEVGTPHGIILDSVDAAFCPGSSRLHIAGR 308
DB 301 LKIGDQVYVYKAGAHFTLLIEGYPKLKEBLN-----IEQVDAAPFCPOQRTAHIIQGRK 355
QY 309 LWWLDLKSQAQTWELPWHHEKVDGALCMKSLGNSGANSAGPYLHGPVLYCYSDV 368
DB 356 IYVINAATPREITLDAAPLPFGDIDAF-----SSDGITIKQGSNYHYHDSF 403
QY 369 EKLNAKALPOPOVNTS-LIGC 389
DB 404 MLVMSRIAPILKVTSAWVC 425

RESULT 4
090WF7 PRELIMINARY; PRT; 439 AA.
ID 090WF7
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Warm-temperature-acclimation-related-65 kDa-protein precursor.
GN WAP65.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
NCBI_TaxID=7962;
RN 1
RP SEQUENCE FROM N.A.
RA Kinoshita S., Ito S., Watabe S.;
RT "cDNA Cloning and Characterization of the Warm-Temperature-
RT Acclimation-Related-65 kDa-Protein."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AB052623; BAB60809.1; -  
 DR InterPro; IPR000585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 5.  
 DR SMART; SM00120; HX; 5.  
 KM SIGNAL  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 439 POTENTIAL.  
 SQ SEQUENCE 439 AA; 50045 MW; 24C9FB96629A9AD CRC64;

Query Match 21.3%; Score 464.5; DB 13; Length 439;  
 Best Local Similarity 29.1%; Pred. No. 1.1e-30;  
 Matches 136; Conservative 69; Mismatches 139; Indels 123; Gaps 18;

QY 15 SLCSLAIATPL-----PTSAHGVABGETKPDPTVERCSGDFATLDDNGT 67  
 DB 6 TLCLALSLFASSDVADDPDTAGIKRPLHHEAKD-----RCA-GMFDIAVNEBIP 59  
 QY 68 LFFKGEFVWKSXKMDRELISERWKNF-----SPVDAFR-----QGHNSVFLIK 112  
 DB 60 YFFKGDHLFKGFHGAELISNE--TFPELDHNNLGHVDAAFRMHSEDSPDHHDQFPFL 116  
 QY 113 GDKVWVPEKKEKGYPLDDEPRGIPSLDAVNECHRGECQAEGLVFPQG----- 164  
 DB 117 DNKVFSTYKHLKLEKQYPRDISDLFPGLDAAVECPKPDCTDVTIFFKGDDELTFPM 176  
 QY 165 -----HGR-----NGTGHNGSTHNGPEX-MRCS 187  
 DB 177 KTKVDEKEFSMNCSTGAFRYMEHYCFHGHQSKPRPVYGDQGRKPKTRIDYFMR 235  
 QY 188 PH-----LVLSALTSNMGATYAFSGTHYRLDTSRGMHSPRIAHQW 230  
 DB 236 PHFGOKSTEEHIEREQCSRVHLDAITSDDCSIVAFRGYHF--VSITGDKFHSPTVESAF 293  
 QY 231 PQGSVAUDAASWEKLYLVGQTQVYVFLTKGTYLVGCGYKLEKEVTHGIIIDSV 290  
 DB 294 KELHSEVDAVFSYSEHLTMDKNEFVYKVBSPHTHLEGYKPKLEVIG-----IBGRVD 348  
 QY 291 AAFICPSSRLHIMAGRLMWLDLKSQAQ-----TWTELPMPHEKLVGALCMESLCPN 345  
 DB 349 AAFVCAHDHIAHVILKGVTVYVDELKATPRAPAKEETIQF-----KTIAMNC-----GPK 399  
 QY 346 SCGANGPGLYLHGNPLVCYSDEKLANAKALPQPNVT-SLGGTTH 391  
 DB 400 -----GVTAVIGNHYLYLDSPKIMMAKIMPEGRHSQGLFGGDH 439

RESULT 5

Q90WR3 PRELIMINARY; PRT; 261 AA.  
 AC 090WR3;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hemopexin (Fragment).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN NCBI\_TaxID=9031;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Weissbach L., Plant P., Grieninger G.;  
 RT "Chicken hemopexin gene expression and sequence of a partial length  
 RT cDNA."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M37319; AA029887.1; -  
 DR InterPro; IPR000585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 3.  
 DR SMART; SM00120; HX; 3.  
 FT NON\_TER 1  
 SQ SEQUENCE 261 AA; 29385 MW; 8C2BBE7010F3C9A0 CRC64;

Query Match 20.9%; Score 456; DB 13; Length 261;  
 Best Local Similarity 36.2%; Pred. No. 2.8e-30;  
 Matches 108; Conservative 46; Mismatches 92; Indels 52; Gaps 11;

QY 96 PVDAAPFOGHNSVFLIKGDKYWTYPPEKKE--KQYPLDDEPRGIPSLDAVECHRG 153  
 DB 10 PCDALAL-WLERYYCLOGTQFYRFRPHSWEVLPQYPRDLRDYF--ICPF----- 55  
 QY 154 CQAEGLVFPQGHGRRNTGGNSGTHHGRPEYMRCSPHLVLSALTSDNMGATYAFSGTHYMR 213  
 DB 56 -----GGRH-----HGN-TSGNACDRCSGE-PFOATISDDSGHIAFRGGLSFR 99  
 QY 214 LDTSRDGMHSPRIAHQWPOGPSAVDAAFSWEKLYLVGQTQVYVFLT-KGQYTLVSGYPK 272  
 DB 100 LDSWRDGMHMPQAHSPGIGQVDAAFSWDKMYTLQSGSVSYVSGRSHQLVEGYPR 159  
 QY 273 RLEKEVSTPHGIIIDSVDAAPICPGSSRLHIMAGRLMWLDLKSQAQATWELPWPEKV 332  
 DB 160 ALQELGV-----KADAAFTCPGSAELYVITGDRMQRVDTLTKSPRNADEPQPLPYDGV 213  
 QY 333 DGLCMESLCPNSCSANGPGLYLHGNPLVCYSDEKLANAKALPQPNVT-TSLGCG 389  
 DB 214 DGAMC-----TADGIYLLRQDRYHRHMDVALLAAPPADPSTAVDLFHC 259

RESULT 6

Q90310 PRELIMINARY; PRT; 445 AA.  
 AC 090310;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE MAM temperature acclimation-related 65-kDa protein precursor.  
 OS Carassius auratus (Goldfish).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 CC Cyprinidae; Carassius.  
 OC NCBI\_TaxID=7957;  
 RN NCBI\_TaxID=7957;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle, and Hepatopancreas;  
 RX MEDLINE=95340486; PubMed=7615502;  
 RA Kikuchi K., Yamashita M., Watabe S., Aida K.;  
 RT "The warm temperature acclimation-related 65-kDa protein, waps5, in  
 RT goldfish and its gene expression."  
 RT J. Biol. Chem. 270:17087-17092 (1995).  
 RN J. Biol. Chem. 270:17087-17092 (1995).  
 RP SEQUENCE OF 1-149 FROM N.A.  
 RC TISSUE=Hepatopancreas;  
 RA Kikuchi K., Watabe S., Aida K.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D50437; BAA08928.1; -  
 DR EMBL; D85428; BAA19835.1; -  
 DR HSSP; P20058; 1HXN.  
 DR InterPro; IPR000585; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 5.  
 DR SMART; SM00120; HX; 4.  
 KM SIGNAL  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 445 WARM TEMPERATURE ACCLIMATION-RELATED 65-  
 FT KDA PROTEIN, WAP5.  
 SQ SEQUENCE 445 AA; 50769 MW; 79C32EAB63BA8266 CRC64;

Query Match 20.5%; Score 447.5; DB 13; Length 445;  
 Best Local Similarity 28.6%; Pred. No. 2.9e-29;  
 Matches 132; Conservative 68; Mismatches 153; Indels 109; Gaps 16;

QY 15 SLCSLAIATPLPTSAHGVABGETKPDPTV-----TRCSGSGFDTATLDDNGTML 68  
 DB 8 TLCLALSLFASSDVADDPDTAGIKRPLHHEAKD-----RCA-GMFDIAVNEBIP 59  
 QY 69 LFFKGEFVWKSXKMDRELISERWKNF-----SPVDAFR-----GHNSVFLIKG 113

```

Db      67 FFKGDHLFKGFHDQALSTNE---TFPELDENHHLGHVDAAFRMSHSDSPAHHDHOFELD 123
Qy      114 DKWVWVPRPEKKEKGYRPLLODEPFGIPSPIDAAVECHRGSCQAEVLFPGQ-----164
Db      124 TKVFYSYKHLKEDYKPDISELFPGLPDHDAVECPTRPCADDTIIFFGDEIYHLDK 193
Qy      165 -----HGH-----NGTGHGNSHTHGPXY-MRCS-188
Db      184 TKVDEKEFESMNCICGAFRMDHYVFRHGHQSKPRPIGVEVQAKPKETRTYFMRCP 242
Qy      189 H-----LVLALTSNDNGATYAFSGTHYWRDLTSDRGMSPIAHQWP 231
Db      243 HFGQKTDEHIEREQCSRVLDAITSDDGSAVAFRGHNP--LSITGDKFHSPTIESAFK 300
Qy      232 QGSSAUDAASFBEKLYLVGQTVYVFLTKGTYLVSGYPRKLEKEVGTGHIILDSYDA 291
Db      301 ELHSEVDVAVSFGHLYMIDNEVFYKVGEPHTEGYPKPEYVIG-----IEGPDVA 355
Qy      292 AFICPGSSRLHIMAGRLMWLDLKSQAQATWTELPWPH-EKVDGALCMEKSLGPNSSCAN 350
Db      356 AFVCAADHIMAHVYKGVTVVDLKAATPRVYVKEGSIHLAKKIDVAMC-----GPK-----405
Qy      351 GPELYLHGNPLYCYSDVEKLNAAKALPDPQNT-SLIGCTH 391
Db      406 --GVTAVIGNHYVQFGSPMIMMAKIMPEQHRVSGLFGCDH 445

RESULT 7
Qy      08JIP9 PRELIMINARY; PRT; 430 AA.
AC      08JIP9;
DT      01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT      01-OCT-2002 (TReMBLrel. 23, Last annotation update)
DE      Warm-temperature-acclimation-related-65 kDa-protein.
GN      MAP65.
OS      Oryzias latipes (Medaka fish) (Japanese ricefish).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC      Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorpha;
OC      Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.
OX      NCBI_TaxID=8092;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=HNT;
RA      Hirayama M., Koshiyama A., Kinoshita S., Watabe S.;
RT      "Primary structural analysis of medaka Map65 and Map65-like protein
RT      and their gene expression in association with temperature acclimation
RT      and during ontogeny.";
RL      Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AB075198; BAB97303.1; -
DR      InterPro: IPR000585; Hemopexin.
DR      Pfam: PF00045; hemopexin; 2.
DR      SMART: SM00120; HX; 4.
SQ      SEQUENCE 430 AA; 49847 MW; A8BFC56B33AAEF15 CRC64;

Query Match      20.4%; Score 445.5; DB 13; Length 430;
Best Local Similarity 28.1%; Pred. No. 4e-29;
Matches 126; Conservative 67; Mismatches 155; Indels 101; Gaps 14;

Qy      16 LCWSIAIATPLPPTSAHGNVAEGETKPPDPTERCSDGMSFDATTLDDNGTMLFFKGEFV 75
Db      10 LCLALVLAIA-----ADHHEHRRKGA-----VRDCK-GIEMDAVAANVEBGIPTFFKEDHL 58
Qy      76 WKSHKMDRELISERWKNP-----SPVDAFR-----QGHNSVFLIKDQKXWVYPRK 123
Db      59 FKGFHGOAELSNKSFALDHDHHLGHVDAAFRMHYEDDLNHDHMFPLDNKVFAYYQHK 118
Qy      124 KEKGYRPLLODEPFGIPSPIDAAVECHRGSCQAEVLFPGQ-----164
Db      119 LEAGYPAISVEVFPGLPDHDAVECPTRPCADDTIIFFGDEIYHLDK 178
Qy      165 -----HGH-----RNGTGHGNSHTHGPXY-MRCS-----187

```

```

Db      179 SMPNCTSAFRMEHNYCFFGHKFSKFPDRTGSEVRGKYPRDAKFFMRCSKFPEDNDHEER 238
Qy      188 ---PHVLALSNDNGATYAFSGTHYWRDLTSDRGMSWPIAHQWPQSPSAVDAFSME 244
Db      239 ERCSRVLHDAITSDAGNIYAFRGHHYIRKDGNDTLKADTIESAKELHSEVDVAFSYN 298
Qy      245 EKLYVQGVVYVFLTKGTYLVSGYPRKLEKEVGTGHIILDSYDAAFICPGSSRLHIM 304
Db      299 SHLYMKDQQLVYRVRGEHHTHLAGYPRKQVQAEIG-----IKGPIDAVCCDRIHAI 353
Qy      305 AGRRLMWLDLKSQAQATWTELPWP-HEKVDGALCMEKSLGPNSSCANSGELYLHGNPLY 363
Db      354 KDRHWYDVMSATPRATNKRPRISILKYVDAMC-----GPG-----GVKVFGRNHYY 401
Qy      364 CYSDVEKLNAAKALPDPQNT-SLIGCTH 391
Db      402 HFESPRTFVAARLPEQHRISLELFGCDH 430

RESULT 8
Qy      P79825 PRELIMINARY; PRT; 446 AA.
AC      P79825;
DT      01-MAY-1997 (TReMBLrel. 03, Created)
DT      01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT      01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE      Hemopexin-like protein (fragment). (Salmo gairdneri).
OS      Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC      Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX      NCBI_TaxID=8022;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=97142135; PubMed=8988369;
RA      Miot S., Duval J., Le Goff P.;
RT      "Molecular cloning of a hemopexin-like cDNA from rainbow trout
RT      liver.";
RL      DNA Seq. 6:311-318 (1996).
DR      EMBL: Z68112; CAA92147.1; -
DR      HSSP: P20058; IHNX.
DR      InterPro: IPR000585; Hemopexin.
DR      Pfam: PF00045; hemopexin; 4.
DR      SMART: SM00120; HX; 4.
FT      NON_TER
SQ      SEQUENCE 446 AA; 50454 MW; 13A8BAEA42B37727 CRC64;

Query Match      20.2%; Score 442; DB 13; Length 446;
Best Local Similarity 27.8%; Pred. No. 8.3e-29;
Matches 126; Conservative 64; Mismatches 168; Indels 96; Gaps 13;

Qy      16 LCWSIAIATPLPPTSAHGNVAEGETKPPD-----VTERCSGMSFDATTLDDNGTMLFF 70
Db      11 LCLVLAISHAHNHAQGCGEDBEGHGHGHLIDRC-QGIBMDAAVAANVEBGIPTFF 69
Qy      71 KGFVWKSHKMDRELISERWKNP-----SPVDAFRQ-----GHSVFLIKDQKXWV 118
Db      70 KGGHVFGRFKAKELSNESFALDHDHHLGHVDAAFRLMHPDCKPTEDHIFFLDTKVP 129
Qy      119 YPEKKEKGYRPLLODEPFGIPSPIDAAVECHRGSCQAEVLFPGQ-----164
Db      130 YHGHQLETFGFPDIDSEVFPGLPDHDAVVCRAPDCEBAVIFFKDELYHYNVTKKYV 169
Qy      165 -----HGH-----RNGTGHGNSHTHGPXY-MRCS-----187
Db      190 EKKEGMPNCTSAFRMEHNYCFFGHQFSKFPDKTGEVGRVPRKEARDVFMKSGKGDYT 249
Qy      188 -----PHVLALSNDNGATYAFSGTHYWRDLTSDRGMSWPIAHQWPQSPSAVDA 239
Db      250 DHIERERCSRVLHDAITSDAGNIYAFRGHHFLEODAGNDVMAADTIBSDFKELHSEVDA 309

```

QY 240 AFSHEKLYVQGTQVYVFLTKGGYTVLSGYPKLEKEVGTPHGIILDSVDAAFCPSGS 299  
DB 310 TFSYENHLVWYKDDKVTYKVGSDHTHDSFKPLKEVIG-----VESEPDAAFVCCDHH 364  
QY 300 RLHMAGRRLMWLULKSQAQATWTELPMP-HEKVDGALCMKESIGPNSCSANGPGLYLIH 358  
DB 365 IAHYIKQTVYDVDLKSPVPVKEGSGTLEFNKDAAMC-----GPE-----GKVLKFK 412  
QY 359 GPNLYCSDVEKLNAAKALPOP-QNVTSLGCTH 391  
DB 413 GNHYFHFQSVKVMMAKAIPEBKTALEFGCDH 446

RESULT 9

Q9DEF1 PRELIMINARY; PRT; 211 AA.  
AC Q9DEF1; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Hemopexin-like protein (Fragment).  
OS Gallitichthys mirabilis (Long-jawed mudsnaker).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidi;  
OC Gobiidae; Giliichthys.  
CX NCBI\_TaxID=8222;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=21117151; PubMed=11172064;  
RA Gracey A.Y., Troll J.V., Somero G.N.;  
RT "Hypoxia-induced gene expression profiling in the euryoxic fish  
RT Giliichthys mirabilis";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).  
DR EMBL; AF286204; AAG13324.1; -.  
DR HSSP; P20058; IHN.  
DR InterPro; IPR000585; Hemopexin.  
DR Pfam; PF00045; hemopexin; 4.  
DR SMART; SM00120; HX; 2.  
FT NON\_TER 1  
SQ SEQUENCE 211 AA; 24104 MW; EC149A2AC07B13CC CRC64;  
Query March 12.4%; Score 271; DB 13; Length 211;  
Best Local Similarity 33.0%; Pred. No. 7.2e-15;  
Matches 69; Conservative 37; Mismatches 83; Indels 20; Gaps 6;  
QY 185 RCSEHLVLSALTSDNHGATVAFSGTHVWRLDTSRDGMHSPDIAHQWPGSPSAVDAFSWE 244  
DB 21 KCS-RVHLDAITSDNAGNMVAFRDHNFIEKDEGDNITVDNIEHAFKELHSDVDVAFSYD 79  
QY 245 EKLVLVGSTQVYVFLTKGGYTVLSGYPKLEKEVGTPHGIILDSVDAAFCPSGSRLHIM 304  
DB 80 DHLTMIKDDHNVKVAEPHHLGYPKTYVEELG-----IDGHVDAFVEDHNIHAFI 134  
QY 305 AGRRLMWLULKSQAQATWTELPMP-HEKVDGALCMKESIGPNSCSANGPGLYLIHGPPLY 363  
DB 135 QGNKIFEVDMKVSREPTEPNERPLALFKVDAAMC-----GPG-----GLKVVGNHFF 182  
QY 364 CYSDVEKLNAAKALPOPQNV-TSLGCTH 391  
DB 183 RFDVLMVAVGRALPEQHRVSTELFGCDH 211

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Proactinopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
CX NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bayne C.J., Gerwick L., Fujiki K., Nakao M., Yano T.;  
RT "Immune-relevant (including acute phase) gene identified in the  
RT livers of rainbow trout, Oncorhynchus mykiss, by means of suppression  
RT subtractive hybridization";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF281339; AAG30014.1; -.  
DR HSSP; P45452; IPEX.  
DR InterPro; IPR000585; Hemopexin.  
DR Pfam; PF00045; hemopexin; 1.  
DR SMART; SM00120; HX; 1.  
FT NON\_TER 1  
SQ SEQUENCE 116 AA; 13379 MW; 4543D625B16ZB3BC CRC64;

RESULT 11

Q9DEF1 PRELIMINARY; PRT; 453 AA.  
AC Q12945; 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Vitronectin.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;  
OC Gallus.  
CX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=White Leghorn; TISSUE=Retina;  
RX MEDLINE=98090066; PubMed=9362471;  
RA Martinez-Morales J.R., Barbas J.A., Marti E., Bovolenta P., Edgar D.,  
RA Rodriguez-Tebar A.;  
RT "Vitronectin is expressed in the ventral region of the neural tube and  
RT promotes the differentiation of motor neurons";  
RL Development 124:5139-5147(1997).  
DR EMBL; Y11030; CA471914.1; -.  
DR HSSP; P45452; IPEX.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR001212; Somatomedin\_B.  
DR Pfam; PF00045; hemopexin; 4.  
DR Pfam; PF01033; Somatomedin B; 1.  
DR PRINTS; PR00022; SOMATOMEDINB.  
DR SMART; SM00120; HX; 4.  
DR SMART; SM00201; SO; 1.  
DR PROSITE; PS00024; HEMOPEXIN; 2.  
DR PROSITE; PS00524; SOMATOMEDIN B; 1.  
SQ SEQUENCE 453 AA; 51636 MW; 82B3BDC2B9397C4 CRC64;  
Query March 9.2%; Score 200.5; DB 13; Length 453;  
Best Local Similarity 25.8%; Pred. No. 1.6e-08;  
Matches 91; Conservative 42; Mismatches 103; Indels 117; Gaps 20;  
QY 26 LPPTSAIGNVAEGSTKP---DPVTERCSGWSFDTATLLDNGTMLPFKGEFVWKS HKWD 82

```

Db      127 LHPTTTTTSDETRNPSLDDEPEELCSRK--PFAFTDLKNGSIVAFRGKYFELDSS 185
Qy      83 -----RELISERKKNFSPVDAAFR-----OGHNSVFLIKGDKWVYPEKKEKGYPLQLQ 133
Db      186 VRDGYPLTISDWM-GIEGPIIDAAFTIRNCG--KTYLFKSSQYWRFDGALDPCYPRDIS 242
Qy      134 DEFGPIPSPLDAAVECHRGEGCOAGVLFPGHGRNGTGNSTHHPPEYWRCSPHLVLS 193
Db      243 EGEGLIPNDIDAA-----FALPAH-----SYGNE-----267
Qy      194 ALTSNDNGATYATPAGTHYWRDLTSRDGWSHWPALAHQWQ-----GPSAV--DAAF---241
Db      268 -----RVYFPKGYKM-----SYDFAHQPTQAECEKSSPSTVFNNYAFWNRD 309
Qy      242 SMEEKLYVQSTGVYVLTGKGYTLVSGYPRKLEK--VETPHGIIIDSDAAFCIGSS 239
Db      310 SWED-----FPLSLFGSRMVGASSQRLISRDMRGVPMQ--LDAMAGRIYVSSR 356
Qy      300 RLHIMAGR-----LM-WLDLKSQAQTWTLPWHEKYDGLC 337
Db      357 QPRRRSRKRKRKYRNHRTNLGLMSWLN--SDSESTDSW-----LSSQC 403

```

## RESULT 12

```

ID      081WC3      PRELIMINARY;      PRT;      634 AA.
AC      081WC3;
DT      01-MAR-2003 (TEMBLrel. 23, Created)
DT      01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DE      01-MAR-2003 (TEMBLrel. 23, Last annotation update)
OS      Hypochemical protein (Fragment).
OC      Homo sapiens (Human).
OC      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN      NCBI_TaxID=9606;
RC      SEQUENCE FROM N.A.
RA      TISSUE=Brain;
RA      Strausberg R.;
RL      Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC040507; AA040507.1; -.
KW      Hypochemical protein.
FT      NON TER
SQ      SEQUENCE 634 AA; 70086 MW; 0D4A78663BAA6AF CRC64;

```

Query Match 9.1%; Score 199; DB 4; Length 634;  
 Best Local Similarity 25.0%; Pred. No. 3.2e-08;  
 Matches 80; Conservative 31; Mismatches 123; Indels 86; Gaps 16;

```

Qy      5 LGAPVALGL-----WSLCWSLAITPL-----PPTSAHGNVAEGETKPPDPVTERCS 51
Db      305 VGDPRLRGLEYEDKRVWQLYGVRESVPTAQPPEPLLPBPDPNKSAPPKQDVPKRS 364
Qy      52 DGMSFDTATLDNGTMLFFKGEFVWKSMDRELISER-----WKNFP--SPVDAAF 101
Db      365 T--HFDVA-QINGEAFEPFKGYKFWRLTR--DRHLVSLQPAQMRFWGLDLHLDSDVAVY 420
Qy      102 -RGHNSVFLIKGDKWVYPEKKEKGYPKLQDEFGIISPLDAAVECHRGEGCOAGVL 160
Db      421 ERTSDHRIVEFKGGRYVWFQDNVNEBEGYPRPVSDFELPPGIDAA-----465
Qy      161 FPGHGRNGTGNSTHHPPEYWRCSPHLVLSLTJSDNGATYATPAGTHYWRDLT--SRD 219
Db      466 FSWAHNDR-----TYFFQDQLYWRDHTRH 491
Qy      220 GHSWPI-AHQWPGPSAVDAAFSWEK-KLYLVQSTGVYVLTGKGYTLVSGYPRKLEK 277
Db      492 MDGYPAGQSPFLMGVSTLUDAMKMSDGAAYFFRG-QEYWKVLDGELEVAQYQSTARD 550
Qy      278 -----VETPHGIIIDSDAA 292
Db      551 WLVCQDSQADGSVAAGVDA 570

```

## RESULT 13

```

ID      099745      PRELIMINARY;      PRT;      240 AA.
AC      099745;
DT      01-MAY-1997 (TEMBLrel. 03, Created)
DT      01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT      01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE      Metalloelastase (Fragment).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN      NCBI_TaxID=9606;
RC      SEQUENCE FROM N.A.
RA      MEDLINE=94043200; PubMed=8226919;
RA      Shapiro S.D., Kobayashi D.K., Ley T.J.;
RT      "Cloning and characterization of a unique elastolytic
RT      metalloproteinase produced by human alveolar macrophages.";
RL      J. Biol. Chem. 268:23824-23829(1993).
RN      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=97187267; PubMed=9034720;
RA      Borden P., Heller R.A.;
RT      "Transcriptional control of matrix metalloproteinases and the tissue
RT      inhibitors of matrix metalloproteinases.";
RL      Crit. Rev. Eukaryot. Gene Expr. 7:159-178(1997).
RN      [2]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=97187267; PubMed=9034720;
RA      Borden P., Heller R.A.;
RT      "Transcriptional control of matrix metalloproteinases and the tissue
RT      inhibitors of matrix metalloproteinases.";
RL      Crit. Rev. Eukaryot. Gene Expr. 7:159-178(1997).
RN      [3]
RN      SEQUENCE FROM N.A.
RA      Lin D., Duncan M., Allen E., Araujo R., Aparicio A., Chai A.,
RA      Chung E., Davis K., Pederspiel N., Hyman R., Kalman S., Komp C.,
RA      Kurd O., Lashkari D., Lew H., Namath A., Oefner P., Roberts D.,
RA      Heller R., Davis R.W.;
RL      "Three Matrix Metalloproteinases on 81kb of Pl insert.";
RL      Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL; U78045; AAB36943.1; -.
DR      HSSP; P39900; LITZ.
DR      InterPro; IPR000585; Hemopexin.
DR      Pfam; PF00045; hemopexin; 4.
DR      SMART; SM00120; HX; 4.
DR      PROSITE; PS00024; HEMOPEXIN; 1.
FT      NON TER
SQ      SEQUENCE 240 AA; 28406 MW; 3E620FAF17EB7CCC CRC64;

```

Query Match 9.0%; Score 196.5; DB 4; Length 240;  
 Best Local Similarity 26.9%; Pred. No. 1.5e-08;  
 Matches 65; Conservative 22; Mismatches 92; Indels 63; Gaps 10;

```

Qy      30 SAHGNVAEGETKPPD--VTERCSDGMSFDTATLDNGTMLFFKGEFVW-----KSHKWRD 83
Db      30 SLYGDKENQRLPNPNKSPALCDPMLSFDAVTVGN-KIFFKDFPVLKYSERKTSV 88
Qy      84 ELISERKKNFSPVDAAFR--OGHNSVFLIKGDKWVYPEKKEKGYPKLQD-EFGIIS 141
Db      89 NLISLMPPLPSGIDAEIARNGVFLFKDQYWLISMLRPPNPKSIHSFGFPNPKY 148
Qy      142 PLDAAVECHRGEGCOAGVLFPGHGRNGTGNSTHHPPEYWRCSPHLVLSLTJSDNG 201
Db      149 KIDAAV-----FNPFRF-----161
Qy      202 ATYAFSGTHYWRDLTSD--GHSWPIAHQWPGPSAVDAAFSWEK-LYLVQSTGVY 256
Db      162 -TYFPVNDQYWRDERQMMDDGYPRLITKNPGIGPK-IDAVFYKANKYYRFGQSNQ 219
Qy      257 VF 258
Db      220 EY 221

```

RESULT 14  
 ID 09TV75 PRELIMINARY; PRT; 464 AA.

```

RA Sorensen K.C., Mitchell B.E.;
RT "Comparison of feline stromelysin-1 (MMP3) cDNA with other known MMP3
RL cDNA sequences.";
DR Submitted (AER-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033627; AAK64605.1; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR006026; NZn_MTPeptide.
DR Pfam; PF00045; hemopexin; 3.
DR Pfam; PF00413; peptidase M10; 1.
DR Pfam; PF03933; peptidase M10_N; 1.
DR PRINTS; PR00138; MATRIXIN.
DR SMART; SM00120; HX; 3.
DR SMART; SM00235; ZmC; 1.
DR PROSITE; PS00546; CYSTEINE SWITCH; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON_TER 1 ..1
SQ SEQUENCE 393 AA; 44093 MW; 148C78EAA6BD4A95 CRC64;

Query Match      8.7%; Score 189.5; DB 6; Length 393;
Best Local Similarity 26.4%; Pred. No. 1.1e-07;
Matches 63; Conservative 23; Mismatches 94; Indels 59; Gaps 8

QY 25 PLPPTSAHGNVAEGET-KPPDVTTERCSDCGSPDATTLDDNGTMLFFKGFEVWVKSHKMDR 83
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 201 PPMWSPEDPAGPTESVPERGRTATCDPALSDFAVS-SLRGEILFFDRHLMRKSLRTR 259
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 84 E-----LLSERWKNPSPSVDAAFRO-GHNSVFLIGDKVMYVPREKKKEGYTKLDD-EFP 137
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 260 EPGFYLMCSFMPSLPSGLDAAYETSVDIYFIKGNDFMANRGTEVOAGYPKGITTLGFP 319
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 138 GIPSPLDAAVECHRGCEQAEGVLFPQGHGHRNLTGTHGNSHTHGPEYMRCSPHLVLSALTS 197
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 320 PTYKKIDAAV-----FD 331
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 198 DNHGATYAFGSTHYWRDLDSRDGMH-SWP--IAHQWPGPSAVDAAFSWEKKLYLVGGT 253
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 332 KEKKKTTFVFYGDKWRFDEKRGQSWEPPGFQIADPPGVDSKVDAAFEAFGFYYPFGNS 390
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Search completed: December 16, 2003, 06:52:11
Job time : 179 secs

```



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 23:41:46 ; Search time 25 seconds  
(without alignments)  
735.498 Million cell updates/sec

Title: US-09-900-448-2

Perfect score: 2185  
Sequence: 1 MARVLGAPVALGLMSLCMSL.....NAKALPQPQVNTSLGCTH 391

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_41\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2139.5	97.9	462	1 HEMO_HUMAN	P02790 homo sapien
2	1700.5	77.8	460	1 HEMO_RABIT	P20058 oxyctolagus
3	1556.5	71.2	460	1 HEMO_RAT	P20059 rattus norv
4	1545.5	70.7	460	1 HEMO_MOUSE	O91x72 mus musculu
5	1544	70.7	459	1 HEMO_PIG	P50828 sus scrofa
6	203	9.3	578	1 MM17_MOUSE	Q97083 mus musculu
7	202.5	9.3	470	1 MM12_HUMAN	P33900 homo sapien
8	199	9.1	606	1 MM17_HUMAN	O91129 homo sapien
9	196	9.0	464	1 MM12_RABIT	P75227 oxyctolagus
10	195.5	8.9	476	1 MM10_MOUSE	O55123 mus musculu
11	195.5	8.9	477	1 MM03_HORSE	Q28397 equus caball
12	189.5	8.7	476	1 MM10_HUMAN	P09238 homo sapien
13	187	8.6	476	1 MM10_RAT	P07152 rattus norv
14	184.5	8.4	468	1 MM01_RABIT	P13943 oxyctolagus
15	184.5	8.4	492	1 MM11_MOUSE	Q02853 mus musculu
16	183.5	8.4	467	1 MM08_HUMAN	P02894 homo sapien
17	181.5	8.3	469	1 MM01_HUMAN	P03956 homo sapien
18	179.5	8.2	488	1 MM11_HUMAN	P24347 homo sapien
19	177.5	8.1	471	1 MM13_RABIT	O62806 oxyctolagus
20	177	8.1	475	1 MM03_RAT	P03957 rattus norv
21	176	8.1	477	1 MM03_HUMAN	P08254 homo sapien
22	176	8.1	471	1 MM13_HORSE	P45452 homo sapien
23	176	8.1	472	1 MM13_HORSE	O19927 equus caball
24	174.5	8.0	469	1 MM01_HORSE	O95825 equus caball
25	173	7.9	477	1 MM03_MOUSE	P28862 mus musculu
26	172	7.9	669	1 MM15_HUMAN	P55111 homo sapien
27	169.5	7.8	477	1 VTNIC_XENLA	O11005 xenopus lae
28	169.5	7.8	478	1 VTNIC_HUMAN	P04004 homo sapien
29	167.5	7.7	508	1 MM19_HUMAN	O99542 homo sapien
30	166.5	7.6	466	1 MM13_RAT	P23097 rattus norv
31	166	7.6	469	1 MM01_PIG	P28863 oxyctolagus
32	166	7.6	478	1 MM03_RABIT	P28863 oxyctolagus
33	165.5	7.6	478	1 VTNIC_MOUSE	P29788 mus musculu

34	165	7.6	657	1 MM15_MOUSE	O54732 mus musculu
35	163	7.5	618	1 MM24_MOUSE	O97082 mus musculu
36	163	7.5	618	1 MM24_RAT	O99066 rattus norv
37	162.5	7.4	645	1 MM24_HUMAN	O945r2 homo sapien
38	161.5	7.4	465	1 MM08_MOUSE	O70138 mus musculu
39	159.5	7.3	466	1 MM08_RAT	O08766 rattus norv
40	158	7.2	459	1 VTNIC_PIG	P48819 sus scrofa
41	155.5	7.1	467	1 MM18_XENLA	O13065 xenopus lae
42	154.5	7.1	472	1 MM13_MOUSE	P33435 mus musculu
43	152	7.0	607	1 MM16_HUMAN	P55152 homo sapien
44	151	6.9	469	1 MM01_BOVIN	P28053 bos taurus
45	151	6.9	607	1 MM16_MOUSE	O9wtr0 mus musculu

## ALIGNMENTS

RESULT 1  
HEMO\_HUMAN STANDARD; PRT; 462 AA.  
ID P02750;  
AC 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hemopexin precursor (Beta-1B-glycoprotein).  
GN HPX.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=88316972; PubMed=2842511;  
RA Altruda F., Poli V., Restagno G., Silengo L.;  
RT "Structure of the human hemopexin gene and evidence for  
RT intron-mediated evolution."  
RL J. Mol. Evol. 27:102-108(1988).  
RN [2]  
RP SEQUENCE OF 2-462 FROM N.A.  
RX MEDLINE=89122012; PubMed=3220477;  
RA Law M.L., Cai G.Y., Hartz J.A., Jones C., Kao F.T.;  
RT "The hemopexin gene maps to the same location as the beta-globin gene  
RT cluster on human chromosome 11."  
RL Genomics 3:48-52(1988).  
RN [3]  
RP SEQUENCE OF 22-462 FROM N.A.  
RX MEDLINE=85242073; PubMed=2989777;  
RA Altruda F., Poli V., Restagno G., Argos P., Cortese R., Silengo L.;  
RT "The primary structure of human hemopexin deduced from cDNA sequence:  
RT evidence for internal, repeating homology."  
RL Nucleic Acids Res. 13:3841-3859(1985).  
RN [4]  
RP SEQUENCE OF ACTIVE PROTEIN.  
RX MEDLINE=85113173; PubMed=3855550;  
RA Takahashi N., Takahashi Y., Putnam F.W.;  
RT "Complete amino acid sequence of human hemopexin, the heme-binding  
RT protein of serum."  
RL Proc. Natl. Acad. Sci. U.S.A. 82:73-77(1985).  
RN [5]  
RP SEQUENCE OF 24-255.  
RX MEDLINE=85076955; PubMed=6510521;  
RA Frantlikova V., Borvak J., Klub I., Moravsek L.;  
RT "Amino acid sequence of the N-terminal region of human hemopexin."  
RL FEBS Lett. 178:213-216(1984).  
RN [6]  
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=84193947; PubMed=6371807;  
RA Takahashi N., Takahashi Y., Putnam F.W.;  
RT "Structure of human hemopexin: O-glycosyl and N-glycosyl sites and  
RT unusual clustering of tryptophan residues."  
RL Proc. Natl. Acad. Sci. U.S.A. 81:2021-2025(1984).  
CC -!- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN  
AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE

CC CIRCULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M36803; AAAS8678.1; -;  
 DR EMBL; M36796; AAAS8678.1; JOINED.  
 DR EMBL; M36799; AAAS8678.1; JOINED.  
 DR EMBL; M36800; AAAS8678.1; JOINED.  
 DR EMBL; M36801; AAAS8678.1; JOINED.  
 DR EMBL; M36802; AAAS8678.1; JOINED.  
 DR EMBL; J03048; AAAS2704.1; -;  
 DR EMBL; X02537; CAA26382.1; ALT\_INIT.  
 DR PIR; I56456; OOHU.  
 DR HSSP; P20058; IHXN.  
 DR SWISS-2DPAGE; P02790; HUMAN.  
 DR Sienna-2DPAGE; P02790; -;  
 DR GeneW; HGNC:5171; HPX.  
 DR MIM; 142290; -;  
 DR GO; GO:0005615; Cytoplasmic space; TAS.  
 DR GO; GO:0005488; F:binding activity; TAS.  
 DR GO; GO:0015232; F:heme transporter activity; TAS.  
 DR GO; GO:0005211; F:plasma glycoprotein; TAS.  
 DR GO; GO:0015886; P:heme transport; TAS.  
 DR GO; GO:0006879; P:iron ion homeostasis; TAS.  
 DR InterPro; IPR005058; Hemopexin.  
 DR Pfam; PF00045; hemopexin; 5.  
 DR SMART; SM00120; HX; 5.  
 DR PROSITE; PS00024; HEMOPEXIN; 2.  
 DR GlycoProtein; Heme; Plasma; Repeat; Transport; Signal.  
 DR KW  
 FT SIGNAL. 1 23  
 FT CHAIN. 24 462  
 FT DOMAIN. 56 93 HEMOPEXIN-LIKE 1.  
 FT DOMAIN. 97 141 HEMOPEXIN-LIKE 2.  
 FT DOMAIN. 188 231 HEMOPEXIN-LIKE 3.  
 FT DOMAIN. 263 306 HEMOPEXIN-LIKE 4.  
 FT DOMAIN. 308 351 HEMOPEXIN-LIKE 5.  
 FT METAL. 79 79 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT METAL. 150 150 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT DISULFID. 50 231  
 FT DISULFID. 149 154  
 FT DISULFID. 188 200  
 FT DISULFID. 257 460  
 FT DISULFID. 366 408  
 FT DISULFID. 418 435  
 FT CARBOHYD. 24 24 O-LINKED.  
 FT CARBOHYD. 64 64 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD. 187 187 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD. 240 240 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD. 246 246 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD. 453 453 N-LINKED (GLCNAC. . .).  
 SQ SEQUENCE. 462 AA; 51676 MW; 054B44D603763B8 CRC64;

Query Match 97.9%; Score 2139.5; DB 1; Length 462;  
 Best Local Similarity. 84.6%; Pred. No. 5.8e-176;  
 Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;

QY 1 MARVLGAPVALGLMSLCWSLAIATPLPTSAHGNAEGETKPPDVTYTERCSQSGMFDATT 60  
 DB 1 MARVLGAPVALGLMSLCWSLAIATPLPTSAHGNAEGETKPPDVTYTERCSQSGMFDATT 60  
 QY 61 LDNGNGMLPFKSGFWKSHKWDRELLSERKKNPSPDPAFRQGHNSVFLIKGDKTWVYP 120  
 DB 61 LDNGNGMLPFKSGFWKSHKWDRELLSERKKNPSPDPAFRQGHNSVFLIKGDKTWVYP 120

QY 121 PEKKEKGYKLLQDERPFGIPSPDLAAVECHRGCEQAGVLPFQ----- 163  
 DB 121 PEKKEKGYKLLQDERPFGIPSPDLAAVECHRGCEQAGVLPFGQDREWFMDLATGTMKER 180  
 QY 164 -----GIGHN 169  
 DB 181 SWPAVNCSSALRWLGRYYCFQGNQFLRPDPVNGEVPYPRDVRDYFMPCCPGRGHGN 240  
 QY 170 GTGHGNSHTHGPEYMCSPHLVLSALTSNDHGATVAFSGTHYWRDLTSDRGHMSPIAHQ 229  
 DB 241 GTGHGNSHTHGPEYMCSPHLVLSALTSNDHGATVAFSGTHYWRDLTSDRGHMSPIAHQ 300  
 QY 230 WPGSPAVDAFAFSEWEKYLVOCTQYVFLTKGTYLVSGYPRLEKEVGTPIHLLDSV 289  
 DB 301 WPGSPAVDAFAFSEWEKYLVOCTQYVFLTKGTYLVSGYPRLEKEVGTPIHLLDSV 360  
 QY 290 DAAFTCPGSSRLHMGRLMWLDLKSQAQATTELPMWHEKVDALCMKESIGPNSCA 349  
 DB 361 DAAFTCPGSSRLHMGRLMWLDLKSQAQATTELPMWHEKVDALCMKESIGPNSCA 420  
 QY 350 NGPGLYLHGPMLCYSDVEKLNAAKALPOPNVTSILGCTH 391  
 DB 421 NGPGLYLHGPMLCYSDVEKLNAAKALPOPNVTSILGCTH 462

RESULT 2  
 HEMO\_RABIT  
 ID HEMO\_RABIT STANDARD, PRT, 460 AA.  
 AC P20058;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hemopexin precursor.  
 GN HPX.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_Taxid=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND HEME-BINDING SITES.  
 RC TISSUE=Liver;  
 RX MEDLINE=93203213; PubMed=7681064;  
 RA Morgan W.T., Master P., Tatum F., Kao S.-M., Alm J., Smith A.;  
 RT "Identification of the histidine residues of hemopexin that  
 coordinate with heme-iron and of a receptor-binding region.";  
 RL J. Biol. Chem. 268:6256-6262(1993).  
 RN [2]  
 RP SEQUENCE OF 26-53.  
 RX MEDLINE=88339942; PubMed=3421961;  
 RA Wellner D., Cheng K.C., Mueller-Eberhard U.;  
 RT "N-terminal amino acid sequences of the hemopexins from chicken, rat  
 and rabbit.";  
 RL Biochem. Biophys. Res. Commun. 155:622-625(1988).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 242-460.  
 RX MEDLINE=96173004; PubMed=8590016;  
 RA Faber H.R., Groom C.R., Baker H.M., Morgan W.T., Smith A., Baker E.N.;  
 RT "1.8-A crystal structure of the C-terminal domain of rabbit serum  
 haemopexin.";  
 RL Structure 3:551-559(1995).  
 CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN  
 CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE  
 CC CIRCULATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.  
 CC -1- SIMILARITY: Contains 5 hemopexin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

Db 389 VDAVERTSDHKIVFPKGDYVWFVKDNNVEGYPREVSD-FSLPFGIDA----- 437  
 QY 156 AEGVLFPQGHNGHNGSTHGHGPEYMCSPHLVLSALTSNKGATYAFSGTHYRLD 215  
 Db 438 ----VFSWANDR-----TYFFKQOLYWRVD 459  
 QY 216 --TSR--DGWHS-WPIAHQWPGPSAVDAFAFWEE-KLYLVQSTGYVFLTKGTYLVSG 269  
 Db 460 DHTRRMDPGYPAGCPFL--WRGVPSMLDDMRKSDGASFFRG-QEYWKVLDSELEAPG 515  
 QY 270 YPKRLEKE--VGTPHGIIILSDVDAFICPGSSRLHIMAGRRLMWLDLKSQAQATWTELP 326  
 Db 516 YPQSTADMDLVCGEP-----LADAEDVGGPQ-----GR-----SGAQ----- 548  
 QY 327 WPEHKVDG-ALCKEKLSPGNSCSANGPGYLIHGPNIYCYSDVEKLNAAKALPQPNVTS 385  
 Db 549 ----DGLAVCS-----CTSDAHL-----ALPSLLLLTP 573  
 QY 386 LL 387  
 Db 574 LL 575  
 RESULT 5  
 AAY8405  
 ID AAY8405 standard; Protein; 587 AA.  
 AC AAY8405;  
 XX 28-JUL-2000 (first entry)  
 XX Mouse MT4-MMP (2) matrix metalloproteinase amino acid sequence.  
 DE Matrix metalloproteinase; MT4-MMP; treatment; arthritis deformans;  
 XX Chronic rheumatoid arthritis; asthma; autoimmune disease; hepatitis;  
 KM atrophic dermatitis; psoriasis; contact dermatitis; hair loss; nephritis;  
 KM ischaemic disease; pancreatitis; arteriosclerosis; leukemia; tumor;  
 KM wound; corneal ulcer; tissue damage; inflammation; cerebral apoplexia;  
 KM Alzheimer's disease; dementia; multiple sclerosis; Parkinson's disease;  
 KM gene therapy.  
 XX Mus sp.  
 OS WO200018900-A1.  
 PN 06-APR-2000.  
 PD 29-SEP-1999; 99WO-JP05349.  
 XX 29-SEP-1998; 98JP-0276258.  
 PR 29-SEP-1998; 98JP-0291505.  
 XX (SEIKI/) SEIKI M.  
 PA Seiki M;  
 XX PI  
 DR MPI: 2000-293141/25.  
 XX N-PSDB; AAA13378.  
 PT Physiologically-active transmembrane matrix metalloproteinase  
 PT polypeptide, useful in screening inhibitors and activators for treating  
 PT e.g. arthritis deformans, asthma and cancers  
 XX  
 PS Claim 11; Page 59-66; 113pp; Japanese.  
 XX This sequence represents the murine matrix metalloproteinase (MT4-MMP 2)  
 CC amino acid sequence. The invention relates to a mammalian transmembrane  
 CC MMP polypeptide, its variants and the DNA sequences encoding them. Also  
 CC included in the invention is a vector comprising the MMP DNA molecules,  
 CC and cells transformed using the vector. The protein, DNA and antibodies  
 CC directed against the polypeptide are useful in screening inhibitors and  
 CC activators for use in treating arthritis deformans, chronic rheumatoid  
 CC arthritis, asthma, autoimmune diseases, atrophic dermatitis, psoriasis,

CC contact dermatitis, hair loss, ischaemic diseases, immune reaction  
 CC accompanying organ transplant, hepatitis, nephritis, pancreaticitis,  
 CC arteriosclerosis, leukaemia, malignant tumours, wounds, corneal ulcers,  
 CC tissue damage or inflammation accompanying leukocytic infiltration,  
 CC together with brain disorders during cerebral apoplexia, Alzheimer's  
 CC disease, dementia, multiple sclerosis, Parkinson's disease or brain  
 CC tumours. The DNA can also be incorporated into a suitable vector for use  
 CC in gene therapy.  
 XX  
 SQ Sequence 587 AA;  
 Query Match 9.3%; Score 204; DB 21; Length 587;  
 Best Local Similarity 24.4%; Pred. No. 3.9e-10;  
 Matches 103; Conservative 34; Mismatches 127; Indels 158; Gaps 25;  
 QY 5 LGAPVALGL-----NSLCSLAIATPL-----PPPSANGNVAEGTKDDPY 46  
 Db 273 VGDPRVYGGPYEDRVRVWQLYGVRESVPTAQLDPEPEBPPLPPPPNNRSSTPQKDV 332  
 QY 47 TERCSQDGSFPAATTLDDNGTMLFPKGFVWKSHKMDRELISER-----WKNP--SP 96  
 Db 333 PNRCTA--HFDAVA-QIRGEAFPFKGFWRLLTR-DRLVSLQPAQMRFRGLPLHDS 388  
 QY 97 VDAAF-ROGHNSVFLIKGDKRWVYPPKEXKKGYPKLLQDEFPGIPSPDLAAVCHRGECQ 155  
 Db 389 VDAVERTSDHKIVFPKGDYVWFVKDNNVEGYPREVSD-FSLPFGIDA----- 437  
 QY 156 AEGVLFPQGHNGHNGSTHGHGPEYMCSPHLVLSALTSNKGATYAFSGTHYRLD 215  
 Db 438 ----VFSWANDR-----TYFFKQOLYWRVD 459  
 QY 216 --TSR--DGWHS-WPIAHQWPGPSAVDAFAFWEE-KLYLVQSTGYVFLTKGTYLVSG 269  
 Db 460 DHTRRMDPGYPAGCPFL--WRGVPSMLDDMRKSDGASFFRG-QEYWKVLDSELEAPG 515  
 QY 270 YPKRLEKE--VGTPHGIIILSDVDAFICPGSSRLHIMAGRRLMWLDLKSQAQATWTELP 326  
 Db 516 YPQSTADMDLVCGEP-----LADAEDVGGPQ-----GR-----SGAQ----- 548  
 QY 327 WPEHKVDG-ALCKEKLSPGNSCSANGPGYLIHGPNIYCYSDVEKLNAAKALPQPNVTS 385  
 Db 549 ----DGLAVCS-----CTSDAHL-----ALPSLLLLTP 573  
 QY 386 LL 387  
 Db 574 LL 575  
 RESULT 6  
 AAE10419  
 ID AAE10419 standard; Protein; 470 AA.  
 AC AAE10419;  
 XX 10-DEC-2001 (first entry)  
 DE Human matrix metalloproteinase-12 (MMP-12) protein.  
 XX Human; matrix metalloproteinase; MMP-12; hair growth; antisense therapy;  
 KM endopeptidase; skin cell; breast cancer; hair follicle; chromosome 11q22.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..16  
 FT Protein /label= Signal\_peptide 17..470  
 FT Domain /label= Mature\_MMP\_12\_protein 90..96  
 FT Domain /label= Cysteine\_switch\_domain 161..185  
 FT Domain /note= "Zinc and calcium binding domain"

XX	Homo sapiens.
XX	WO200190366-A2.
PD	
XX	29-NOV-2001.
PF	
XX	24-MAY-2001; 2001WO-US17076.
PR	
XX	24-MAY-2000; 2000US-206690P.
PA	(CUPRA-) CUPAGEN CORP.
PI	
DR	Leach MD, Shinkets RA;
XX	
XX	WPI; 2002-106200/14.
XX	N-PBDB; ABN75603.
PT	
PT	Novel human polypeptides and polynucleotides useful for diagnosing,
PT	preventing and treating cardiovascular disease, neurodegenerative,
PT	hyperproliferative disorders and disorders related to organ
PT	transplantation -
XX	
PS	Claim 10; Page 541-542; 2508pp; English.
XX	
CC	Sequences ABP10208-ABP35561 represent 4534 novel human proteins
CC	designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC	ABN79387 represent CDNA's encoding them. The invention also encompasses
CC	polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC	referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC	the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC	polynucleotides, the recombinant production of ORFX proteins, antibodies
CC	specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC	polypeptides, methods of screening for modulators of ORFX expression or
CC	activity, and methods of screening individuals for a predisposition to an
CC	ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC	range of biological activities, such as cytokine, cell proliferation,
CC	cell differentiation, immune modulation, haematopoiesis regulation,
CC	tissue growth, angiogenesis, activin or inhibin activity, chemotactic/ chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, antiinflammatory activity, tumour inhibition activity, CC and antifibrotic activity, and may also be involved in the determination CC of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases. XX XX
SQ	Sequence 87 AA:
Query Match	11.8%; Score 258; DB 23; Length 87;
Best Local Similarity	86.5%; Pred. No. 2e-16;
Matches 45; Conservative	3; Mismatches 4; Indels 0; Gaps 0
Gy	114 DKWWYPPBEKKKGYPKXLDGFFRIPSPDAVAVCHRGECQAEGYLFRGGH 165
Db	5 DKWWYPPBEKKENGYPKLFQEEFPPIPPDDAAVBCHRGECOSBGVLFRGGN 56

XX		AA90502;	
AC			
XX	DT	15-AUG-2000 (first entry)	
XX			
XX			
DE		Murine transmembrane matrix metalloprotease MT4-MMP (2).	
XX			
KW		Matrix metalloprotease; MT4-MMP(2); transmembrane; mouse; antibody;	
KV		drug screening; diagnosis; arthropis deformans; rheumatoid arthritis;	
KW		asthma; autoimmune disease; atopic dermatitis; skin disorder;	
KV		ischæmic disorder; arteriosclerosis; cancer; tissue damage;	
KW		inflammatory disorder; neurological disorder.	
XX			
OS		Mus sp.	
XX			
FN		WO20001805-A1.	
XX			
PD		06-APR-2000.	
XX			
PE		29-SEP-1999; 99WO-JP05350.	
XX			
PR		29-SEP-1998; 98JP-0291501.	
PR		29-SEP-1998; 98JP-0291503.	
XX			
PA		(KXOW ) KYOMA HAKKO KOGYO KK.	
P1		Hanai N, Furuya A;	
DR		WPI; 2000-293115/25.	
XX		N-PADB; AAA14339.	
PT			
PS		Claim 1; Page 86-90; 152pp; Japanese.	
XX			
CC		This sequence represents a novel murine transmembrane matrix	
CC		metalloprotease MT4-MMP (2). MT4-MMP (2) has physiological activity	
CC		different to that of MT4-MMP. The invention relates to antibodies	
CC		against both murine and human MT4-MMP (2) (AA90502 and AA90503) and to	
CC		antibodies against a novel murine and human transmembrane matrix	
CC		metalloprotease MTS-MMP (AA90504 and AA90505). The antibodies are	
CC		useful for the diagnosis and screening of inhibitors and activators	
CC		such as arthropis deformans, rheumatoid arthritis, asthma, autoimmune	
CC		diseases and atopic dermatitis. The anti-MTS-MMP antibodies may	
CC		additionally be used to diagnose and screen therapeutic or prophylactic	
CC		agents for conditions affecting the brain, such as cerebral stroke and	
CC		Alzheimer's disease. The antibodies of the invention may also be used for	
CC		the prevention, diagnosis and treatment of psoriasis, contact	
CC		dermatitis, hair loss, ischaemic diseases, immune reaction accompanying	
CC		organ transplant, hepatitis, nephritis, pancreatitis, arteriosclerosis,	
CC		leukaemia, malignant tumours, wounds, corneal ulcers, tissue damage or	
CC		inflammation accompanying leukocytic infiltration, dementia, multiple	
CC		sclerosis, Parkinson's disease or brain tumour.	
SQ			
	Sequence	587 AA;	
	Query Match	9.3%; Score 204; DB 21; Length 587;	
	Best Local Similarity	24.4%; Pred. No. 3.9e-10;	
	Matches 103; Conservative	34; Mismatches 127; Indels 158; Gaps 25	
Dd			
	5 LGAPVALGL-----WSLCWLSLAIRPL-----PRTSAGNVAGETKPPDDV 46		
	273 VGDFVRKGLPFEDRYVKWKLYGVRESSPTAQDLTPREBEERPLLPERPNRSSPTPOXDY 332		
Oy			
	47 TERCSDMSFPDATTLDNGTMLPFKGGEFVMKSHKMBELISER-----WKNP--SP 96		
	333 PHRCTA--HFDVAH-QLRGEAFPFKKGYPMRLTR-DRLVLVSQRAQMHRFMRGRLPHDS 388		
Oy			
	97 VDAAF-RQGNSVFLIGDKYKWTVPRPEKEKGYPKLLODEFPGLIPSLDAAVECHRGECQ 155		

XX WPI: 2001-476164/51.  
 DR N-PSDB; AAB98592.  
 XX  
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 antibodies and research use -  
 XX  
 PS Claim 20; Page 1017-1018; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostic,  
 CC forensic, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 SQ Sequence 462 AA;  
 Query Match 97.9%; Score 2139.5; DB 22; Length 462;  
 Best Local Similarity 84.6%; Pred. No. 8.9e-193;  
 Matches 391; Conservative 0; Mismatches 0; Indels 71; Gaps 1;  
 QY 1 MARVLGAPVALGLMSLCWSLAIAATPLPPTSAHGNVAGETKPPDVTERTSCDSGMSPDATT 60  
 DB 1 MARVLGAPVALGLMSLCWSLAIAATPLPPTSAHGNVAGETKPPDVTERTSCDSGMSPDATT 60  
 QY 61 LDDNGTMLPFKGEFVWVKSHKMDRELISERWKNPSPVDAAFROGHSNVPFLIKDKVWVVP 120  
 DB 61 LDDNGTMLPFKGEFVWVKSHKMDRELISERWKNPSPVDAAFROGHSNVPFLIKDKVWVVP 120  
 QY 121 PERKEKGYPLTLDPEFGIPSPIDAIVECHRGECQACGVI-PQ----- 163  
 DB 121 PERKEKGYPLTLDPEFGIPSPIDAIVECHRGECQACGVI-PQ----- 163  
 QY 164 -----GHGHRN 169  
 DB 181 SWPVGNCSSALMLGRIYFPGNQPLRPDPVGEVPRYPDVRDYFMPCPGRGHNRN 240  
 QY 170 GTGHGNSHTHGPEYMRCSPLVLALTSDNHGATYAFSGTHYRLDTSRDGMSWPIAHQ 229  
 DB 241 GTGHGNSHTHGPEYMRCSPLVLALTSDNHGATYAFSGTHYRLDTSRDGMSWPIAHQ 300  
 QY 230 WPGGPSAVDAAFSWEKLYIVOGTQVYVFLTKGTYTVSGYPRKLEKEVGTPIGIIIDSV 289  
 DB 301 WPGGPSAVDAAFSWEKLYIVOGTQVYVFLTKGTYTVSGYPRKLEKEVGTPIGIIIDSV 360  
 QY 290 DAAPFCGSSRLHIMAGRLMWLIDLKSGAQTWTELPMPEKVDGALCMKESIGPNSCSA 349  
 DB 361 DAAPFCGSSRLHIMAGRLMWLIDLKSGAQTWTELPMPEKVDGALCMKESIGPNSCSA 420  
 QY 350 NGPGLYLHGPNTLYCYSDEKLNAAKALPOPONTSLAGCTH 391  
 DB 421 NGPGLYLHGPNTLYCYSDEKLNAAKALPOPONTSLAGCTH 462  
 RESULT 2  
 AAG00304  
 ID AAG00304 standard; Protein; 83 AA.  
 XX  
 AC AAG00304;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 4385.  
 XX  
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.

XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-020610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GENT) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 XX  
 DR N-PSDB; AAC00310.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 4385; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 83 AA;  
 Query Match 20.9%; Score 456; DB 21; Length 83;  
 Best Local Similarity 98.8%; Pred. No. 3.9e-35;  
 Matches 82; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MARVLGAPVALGLMSLCWSLAIAATPLPPTSAHGNVAGETKPPDVTERTSCDSGMSPDATT 60  
 DB 1 MARVLGAPVALGLMSLCWSLAIAATPLPPTSAHGNVAGETKPPDVTERTSCDSGMSPDATT 60  
 QY 61 LDDNGTMLPFKGEFVWVKSHKMDR 83  
 DB 61 LDDNGTMLPFKGEFVWVKSHKMDR 83  
 RESULT 3  
 ABP31577  
 ID ABP31577 standard; Protein; 87 AA.  
 XX  
 AC ABP31577;  
 XX  
 DT 09-JUL-2002 (first entry)  
 XX  
 DE Human glycoprotein-like ORF550 protein, SEQ ID NO:1100.  
 XX  
 KM Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KM disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KM immune modulation; haematopoiesis regulation; tissue growth; haemostatic;  
 KM angiogenesis; activin; inhibitor; chemotactic; chemokinetic; fertility;  
 KM thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KM behaviour; cancer; proliferative disorder; neurological disorder;  
 KM cardiovascular disease; immune system disorder; organ transplantation;  
 KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KM hypochyroidism; cholesterol ester storage disease; infection; vulnery;  
 KM vasotropic; antiporatic; antidiabetic; cytostatic; neurotropic;  
 KM neuroprotective; antilatherosclerotic; anticoncugant; thrombolytic;  
 KM cardiant; hypotensive; antichyroid; antiinflammatory; immunomodulator;  
 KM dermatological; analgesic; virucide; antibacterial; fungicide.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:37:44 ; Search time 64 Seconds  
(without alignments)  
969.720 Million cell updates/sec

Title: US-09-900-448-2  
Perfect score: 2185  
Sequence: 1 MARVLGAPVALGLMISLCWEL.....NAKALPQPNTSLIGCTH 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A.Geneseq.15Jun03.\*  
2: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2139.5	97.9	462	22	AAAM23933
2	456	20.9	83	21	AAAG00304
3	258	11.8	87	23	ABP31577
4	204	9.3	587	21	AAAY90502
5	204	9.3	587	21	AAAB8405
6	202.5	9.3	470	22	AAAI0419
7	202.5	9.3	470	22	AAAB84614
8	202.5	9.3	470	22	AAAB74595
9	202.5	9.3	470	22	AAAB49982

10	202.5	9.3	470	23	AAU91061	Human head and nec
11	202.5	9.3	470	24	ABP96800	Human COPD related
12	202.5	9.3	470	24	ABR48203	Human bladder can
13	202.5	9.3	470	24	ABU56421	Lung cancer-associ
14	202.5	9.3	470	24	ABU56663	Lung cancer-associ
15	202.5	9.3	473	21	AAAB3772	Human cancer-associ
16	199	9.1	519	22	AAE10426	Human matrix metal
17	199	9.1	519	22	AAAB84619	Amino acid sequenc
18	199	9.1	605	21	AAAY90503	Human transmembran
19	199	9.1	605	21	AAO19713	MT4-MMP catalytic
20	199	9.1	606	21	AAV88404	Human MT4-MMP (2)
21	196	9.0	464	19	AAW52135	Rabbit matrix meta
22	195.5	8.9	477	23	AAU91062	Human head and nec
23	192.5	8.8	476	24	ABU03460	Angiogenesis-associ
24	189.5	8.7	475	23	ABU05574	Breast cancer-asso
25	189.5	8.7	475	24	ABU56595	Lung cancer-associ
26	189.5	8.7	476	22	AAE10421	Human matrix metal
27	189.5	8.7	476	24	ABR48412	Amino acid sequenc
28	189.5	8.7	476	24	ABR48147	Human bladder can
29	189.5	8.7	476	24	ABP54455	Matrix metalloprot
30	183.5	8.4	444	22	AAAG5357	Human MMP-8alt pol
31	183.5	8.4	467	22	AAE10416	Human matrix metal
32	183.5	8.4	467	22	AAAB84610	Amino acid sequenc
33	183.5	8.4	467	22	AAAG6358	Human neutrophil c
34	181.5	8.3	469	22	AAE10415	Human matrix metal
35	181.5	8.3	469	22	AAAB84606	Amino acid sequenc
36	181.5	8.3	469	24	ABR48148	Human bladder can
37	181.5	8.3	469	24	ABU56596	Lung cancer-associ
38	181.5	8.3	469	24	ABU56597	Lung cancer-associ
39	181.5	8.3	469	24	ABU07454	Protein different
40	181.5	8.3	469	24	ABP54454	Matrix metalloprot
41	181.5	8.3	492	13	AAAR24863	Sequence of pre-pr
42	181.5	8.3	496	22	AAAG7509	Human colon cancer
43	180.5	8.3	457	10	AAAP93628	Sequence of human
44	180.5	8.3	469	8	AAAT70611	Sequence encoded b
45	179.5	8.2	479	23	ABP41861	Human ovarian anti

## ALIGNMENTS

RESULT 1	AAAM23933	standard; Protein; 462 AA.
ID	AAAM23933;	
XX	12-OCT-2001	(first entry)
XX	Human EST encoded protein SEQ ID NO: 1458.	
DE	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;	
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;	
KW	diagnostics; forensic test; gene mapping; genetic disorder;	
KW	biodiversity; gene therapy; nutrition.	
OS	Homo sapiens.	
PN	WO200154477-A2.	
XX	02-AUG-2001.	
PF	25-JAN-2001; 2001WO-US02687.	
XX	25-JAN-2000; 2000US-0491404.	
PR	17-JUL-2000; 2000US-0617746.	
PR	03-AUG-2000; 2000US-0631451.	
PR	15-SEP-2000; 2000US-0663870.	
XX	(HYSE-) HYSEQ INC.	
PA	Tang YF, Liu C, Zhou F, Qian XB, Wang Z, Chen R, Asundi V;	
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;	



```
Db      548 GCTATACCCAGTAAGGNTTATCCGAGGGGCTGCAGAAAGTAGAGACCCCTCATG 489
QY      862 GGAATTATCTGGAAGCTGTGGAATGCGGCTTTATCTGCGCTTCTCGGCTCCATA 921
Db      488 GGAATTATCTGTAATCTGTGGAATGCGGCTTTATCTGCGCTTCTCGGCTCCATA 430
QY      922 TCATGGCAGGACGGGCGCTGTGGTGTGAGACCTGAAGTACAGAGCCCAAGCCACGNGA 981
Db      429 TCAATGGCAGGACGGGCGCTGTGGTGTGAGACCTGAAGTACAGAGCCCAAGCCACGNGA 370
QY      982 CAGAGCTTCCCTTGGCCCCCATGAGAGGTAGACGAGGCTTGTGTATGAAAAATGCCCTTG 1041
Db      369 CAGAGCTTCCCTTGGCCCCCATGAGAGGTAGACGAGGCTTGTGTATGAAAAATGCCCTTG 310
QY      1042 GCCCTTAATCTATGTTCCGCCAATGATGATCCGCTTGTACTCATTCATGATGATCCCAATTGT 1101
Db      309 GCCCTTAATCTATGTTCCGCCAATGATGATGATCCGCTTGTACTCATTCATGATGATCCCAATTGT 250
QY      1102 ACTGCTACAGTGAATGTGAGAACTGAAATGACAGCCAGGCTTCCGCCAATCCCAAGATG 1161
Db      249 ACTGCTACAGTGAATGTGAGAACTGAAATGACAGCCAGGCTTCCGCCAATCCCAAGATG 190
QY      1162 TGAACAAGTCTCTGGGCTGCACTCACTGAGGAGGCTTCTGACATGAGTCTGAGGCTG 1221
Db      189 TGAACAAGTCTCTGGGCTGCACTCACTGAGGAGGCTTCTGACATGAGTCTGAGGCTG 130
QY      1222 CACTTCTTAATGTTCTCTAATAATAAGACAGATTGCTTCTTGGCTTCTGACTGAGGAGGCTT 1281
Db      129 CACTTCTTAATGTTCTCTAATAATAAGACAGATTGCTTCTTGGCTTCTGACTGAGGAGGCTT 70
QY      1282 CTGACATGAGTCTGGGCTGAGCCCACTGCCAGTTTCTCATATAATAAGACAGATTGCTT 1341
Db      69 CTGACATGAGTCTGGGCTGAGCCCACTGCCAGTTTCTCATATAATAAGACAGATTGCTT 10
QY      1342 CTTCACTTG 1350
Db      9 CTTCACTTG 1
```

Search completed: December 16, 2003, 05:33:59  
Job time : 3998.17 secs

RESULT	14
AV653336	631 bp mRNA linear EST 15-JAN-2002
LOCUS	AV653336 GLC Homo sapiens cDNA clone GDCUHJ02 3', mRNA sequence.
ACCESSION	AV653336
VERSION	AV653336.1 GI:98744350
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Bukacynski, Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 631)
AUTHORS	Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Pu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z., and Han, Z.
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
PUBMED	21625106
COMMENT	11752456 Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai. Location/Qualifiers 1..631 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="GDCUHJ02" /feature="GLC"UHO2" /tissue_type="corresponding non cancerous liver tissue" /dev_stage="Adult" /lab_host="SOLR" /clone_idb="GLC" /note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
FEATURES	source
BASE COUNT	138 a 180 c 161 g 149 t 3 others
ORIGIN	
Query Match	16.7%; Score 530.6; DB 9; Length 631;
Best Local Similarity	98.7%; Pred. NO. 6.5e-148;
Matches	533; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Dy	823 ATCCGAAGCGGCTGGAGAAGAAAGTCGGAGACCCTCATGGATTAATCTCGAATTCTGTGG 882
Db	1 ATCCGAAGCGGCTGGAGAAGAAAGTCGGAGACCCTCATGGATTAATCTCGAATTCTGTGG 60
Dy	883 ATGGGGCTTTATATCTGCCCTGGGTTTTTCTGGCTTCATATATGAGCAAGAGCGGCTGT 942
Db	61 ATGGGGCTTTATATCTGCCCTGGGTTTTTCTGGCTTCATATATGAGCAAGAGCGGCTGT 120
Dy	943 GTGGGCTGACTGAAGTCAGAGAGCCCAAGCCACGTGACAGAGCTTCTTGCCCCCATG 1002
Db	121 GTGGGCTGACTGAAGTCAGAGAGCCCAAGCCACGTGACAGAGCTTCTTGCCCCCATG 180
Dy	1003 AGAAGTAGACGAGACCTTGTTGTATGAAAAAGTCCCTTGAGCCCTTAATCATGTTCCGCA 1062
Db	181 AGAAGTAGACGAGACCTTGTTGTATGAAAAAGTCCCTTGAGCCCTTAATCATGTTCCGCA 240
Dy	1063 ATGGTCCCGGCTTGATCTCATTCATGGTCCCAATTTGTACTGTACAAGTATGTGAGA 1122
Db	241 ATGGTCCCGGCTTGATCTCATTCATGGTCCCAATTTGTACTGTACAAGTATGTGAGA 300
Dy	1123 AACTGAATGACAGCCAAAGCCCTTCGCGAACCCAGAAATGTACAGCATGTCCTCGGGCTGCA 1182
Db	301 AACTGAATGACAGCCAAAGCCCTTCGCGAACCCAGAAATGTACAGCATGTCCTCGGGCTGCA 360

QY	1183	CTCAGTACGGGGCCCTTCTGACATAGTGTGGCGTCGGCCCGACCTTCCTGTTCTTCATAT	1244
Db	361	CTCACTAGGGGGCCCTTCTGACATAGTGTGGCGTCGGCCCGACCTTCCTGTTCTTCATAT	420
QY	1243	AAAGCAGATTGCTTCTTCCTTCCTCCTCAGTGGGGCCCTTCTGACATAGTGTGGCGTGGC	1302
Db	421	AAAGCAGATTGCTTCTTCCTTCCTCCTCAGTGGGGCCCTTCTGACATAGTGTGGCGTGGC	480
QY	1303	CCCACTTCCCAAGTTTCTCATATTAAGACAGATTGCTTCTTCCTCCTGATCAAGGAC	1362
Db	481	CCCACTTCCCAAGTTTCTCATATTAAGACAGATTGCTTCTTCCTCCTGATCAAGGAC	540
RESULT 15			
AA705094/c			
LOCUS			
DEFINITION	AA705094	548 bp	mrna
ACCESSION	z133h10.61	Soares fetal liver spleen INFLS S1	Homo sapiens CDNA
VERSION	clone IMAGE:462499	3'	similar to gb:J03048 HEMOPLEXIN PRECURSOR
KEYWORDS	(HMAN);	mRNA sequence.	
SOURCE	AA705094		
ORGANISM	AA705094.1	GI:2715012	
	EST.		
	Homo sapiens		
	Homo sapiens (human)		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1	(bases 1 to 548)	
AUTHORS	Hillier, L., Allan, M., Bowles, L., Dubuque, T., Geisel, G., Joet, S.,		
	Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin		
	, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,		
	White, Y., Wyllie, T., Waterston, R. and Wilson, R.		
TITLE	WashU-NCI human EST Project		
COMMENT	Unpublished		
	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LMLT; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Seq primer: -40m13 fwd. RT from Amersham		
	High quality sequence stop: 462.		
FEATURES			
Source	Location/Qualifiers		
	1..548		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="GDB:3753522"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:462499"		
	/sex="male"		
	/dev_stage="20 week-post conception fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/clone_1lb="Soares fetal liver spleen INFLS S1"		
	/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)		
	with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;		
	This is a subcloned version of the original Soares fetal		
	liver spleen INFLS library. 1st strand cDNA was primed		
	with a Pac I - clligo (dr) primer 15'		
	AACGGAAGATTAATTAAGATCTTTTTTTTTTTTTTTT 31',		
	double-stranded cDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Pac I and cloned into the Pac I		
	and Eco RI sites of the modified pT73 vector. Library		
	went through one round of normalization. Library		
	constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	137 a	137 c	160 g
ORIGIN			
Query Match	16.6%	Score 529.6	DB 9; Length 548;
Best Local Similarity	98.9%	Pred. No. 1.2e-147	
Matches	543; Conservative	0; Mismatches	5; Indels
			1; Gaps
			1;

BASE COUNT 123 a 167 c 169 g 143 t  
ORIGIN

Query Match 16.7%; Score 533.2; DB 9; Length 602;  
Best Local Similarity 91.3%; Pred. No. 1e-148;  
Matches 599; Conservative 0; Mismatches 3; Indels 54; Gaps 1;

QY 618 GCCACCTATGCTTCACTAGTGGGACCACTACTAGGCTGTGACACACGACCGGAGATGGCTGG 677  
DB 1 GCCACCTATGCTTCACTAGTGGGACCACTACTAGGCTGTGACACACGACCGGAGATGGCTGG 60  
QY 678 CATAGCTGGCCCATTTGCTCATCACTAGTGGGACCGGCTTTCAAGCATGATGCTGCTTT 727  
DB 61 CATAGCTGGCCCATTTGCTCATCACTAGTGGGACCGGCTTTCAAGCATGATGCTGCTTT 120  
QY 738 TCCGAGGAGAAAACCTATCTGCTGACAGGACCCAGGATATGCTTCTTCTGACAAAG 797  
DB 121 TCCGAGGAGAAAACCTATCTGCTGACAGGACCCAGGATATGCTTCTTCTGACAAAG 180  
QY 798 GGAAGCTATACCTAGTAAAGCGGTTATCCGAAAGCGGCTGGAAGAAAGATCGGACCCCT 857  
DB 181 GGAAGCTATACCTAGTAAAGCGGTTATCCGAAAGCGGCTGGAAGAAAGATCGGACCCCT 240  
QY 858 CATGGATTATCTGATCTGATGATGCGGCTTTATCTGCGGCTTTCTCGGCTC 917  
DB 241 CATGGATTATCTGATCTGATGATGCGGCTTTATCTGCGGCTTTCTCGGCTC 300  
QY 918 CATATCATGACAGAGACGGGCTGTGCTGCTGACCTGAAAGTCAAGAGCCCAAGCCAG 977  
DB 301 CATATCATGACAGAGACGGGCTGTGCTGCTGACCTGAAAGTCAAGAGCCCAAGCCAG 360  
QY 978 TGAACAGAGCTCTTGGGCGCCCATGAGAAAGTGAACGAGCTTGTATGAGAAAGTCC 1037  
DB 361 TGAACAGAGCTCTTGGGCGCCCATGAGAAAGTGAACGAGCTTGTATGAG----- 412  
QY 1038 CTGGGCTTAACTGATGTTCCGCAATGTCGCGGCTTGTACTCATGATGTCCTCAAT 1097  
DB 413 -----TTCATGATGCCAAT 426  
QY 1098 TTGACTGCTACAGTGAATGAGAAACCTGAATGACACCAAGCCCTTCGCAACCCAG 1157  
DB 427 TTGACTGCTACAGTGAATGAGAAACCTGAATGACACCAAGCCCTTCGCAACCCAG 486  
QY 1158 AATGTGACAGCTCTTGGGCTGCTGACCTGACAGGCGGCTTGTGACATGAGTCTGAGCTG 1217  
DB 487 AATGTGACAGCTCTTGGGCTGCTGACCTGACAGGCGGCTTGTGACATGAGTCTGAGCTG 546  
QY 1218 GCCCACTCTGATGTTCTCATATAAAGACAGATTGCTTCTTCTGCTTCTCACTGA 1273  
DB 547 GCCCACTCTGATGTTCTCATATAAAGACAGATTGCTTCTTCTGCTTCTCACTGA 602

RESULT 13  
BX412777/c 914 bp mRNA linear EST 13-MAY-2003  
LOCUS BX412777 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
DEFINITION CS0DP010YK01.3-PRIME, mRNA sequence.  
ACCESSION BX412777  
VERSION BX412777.1 GI:30649532  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1. (bases 1 to 914) Jesse, J. and Polayes, D.  
AUTHORS L., W. B., Gruber, C.;  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAK031CB08NM1cluster=5958.r)  
Feng Liang Email: [liang@lifetech.com](mailto:liang@lifetech.com) URL :  
[http://fulllength.invitrogen.com/](http://fulllength.invitrogen.com/InvitrogenCorporation1600)  
Faraday Avenue Genoscope sequence ID : CS0BAK031CB08NM1.

#### FEATURES

source  
1. 914  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DP010YK01"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

BASE COUNT 182 a 275 c 199 g 258 t  
ORIGIN

Query Match 16.7%; Score 531.2; DB 13; Length 914;  
Best Local Similarity 93.4%; Pred. No. 5.9e-148;  
Matches 577; Conservative 0; Mismatches 38; Indels 3; Gaps 2;

QY 722 AGTGAATGCTGCTTCTTCTGAGAGAAAAA-CCTATCTGCTCAAGGACCCAGGTA 779  
DB 629 AGTGAATGCTGCTTCTTCTGAGAGAAAAAAGGTTATCTGGAACGGGCGCCAGGTA 570  
QY 780 TATGCTTCTGCAAGAGAGAGCTTATACCTTA-GRAGCGGTTATCCAGAGGCTGGA 838  
DB 569 TATGATTCCTGCAAGAGAGAAAAATATACCTTAAGTTAAGTTGCGAAAGGCTGGA 510  
QY 839 GAAGAGAGTGGGACCCCTCATAGGATTAATCTGCACTCTGAGATGCGGCTTTATCTG 898  
DB 509 GAAGAGAGTGGGACCCCTCATAGGATTAATCTGCACTCTGAGATGCGGCTTTATCTG 450  
QY 449 CCATGGGTATATGCGGCTCCATATCATGACAGAGCGGCTGTGGCTGGAACCTGAA 390  
DB 899 CCTGGGTCTTCTGCGCTTCATATCATGAGAGAGCGGCTGTGGCTGGAACCTGAA 958  
QY 959 GTCAGAGAGCCCAAGCCAGAGTGAACAGGCTTCTTGGCCCATGAGAAAGGTAGACGAGC 1018  
DB 389 GTCAGAGAGCCCAAGCCAGAGTGAACAGGCTTCTTGGCCCATGAGAAAGGTAGACGAGC 330  
QY 1019 CTGTGTATGAGAAAAATGCTTGGGCTTGAATCATGATGTCGCAATGATCCGGCTTGA 1078  
DB 329 CTGTGTATGAGAAAAATGCTTGGGCTTGAATCATGATGTCGCAATGATCCGGCTTGA 270  
QY 1079 CCTGATCATGATGTCGCAATGTCATGCTACAGTGAATGAGAGAACTGAATGACCCAA 1138  
DB 269 CCTGATCATGATGTCGCAATGTCATGCTACAGTGAATGAGAGAACTGAATGACCCAA 210  
QY 1139 GGCCCTTCCGCAACCCCAAGATGTAACAGTCTTCCGCTGCTGACCTGAGAGGCGCTT 1198  
DB 209 GGCCCTTCCGCAACCCCAAGATGTAACAGTCTTCCGCTGCTGACCTGAGAGGCGCTT 150  
QY 1199 CTGACATGATGTCGCTGCGCCCACTCTCTAGTCTCTCATATAAAGACAGATTGCTTC 1258  
DB 149 CTGACATGATGTCGCTGCGCCCACTCTCTAGTCTCTCATATAAAGACAGATTGCTTC 90  
QY 1259 TTGCTTCTCACTGAGAGGCGCTTCTGACATGATGTCGCTGCGCCCACTCTCCAGTTT 1318  
DB 89 TTGCTTCTCACTGAGAGGCGCTTCTGACATGATGTCGCTGCGCCCACTCTCCAGTTT 30  
QY 1319 CTCATATAAAGACAGAT 1336  
DB 29 CTCATATAAAGACAGAT 12

```

Db      905 CTTGAGTGGAGCCCACTACTGCGCTGTGACACACCGGAGTGGCTGAGTACGTGCGC 964
QY      689 CATTGCTATCAGTGGTCCCGAGGGTCTTCAGCAGTGTGATGCTCTTTCTCGGGAAGA 748
Db      965 CATTGCTATCAGTGGTCCCGAGGGTCTTCAGCAGTGTGATGCTCTTTCTCGGGAAG- 1023
QY      749 AAAACTCTATCTGTGTCCAGGGCACC 773
Db      1024 AAAACTCTATCTGTGTCCAGGGCACC 1048

RESULT 11
LOCUS   A1798878      562 bp      mRNA      linear      EST 18-DEC-1999
DEFINITION we93g08.x1 Soares NFL T GBC_S1 Homo sapiens cDNA clone
IMAGE:2348702.3' similar to gb:U03048 HEMOPEXIN PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION A1798878      GI:5364350
VERSION   A1798878
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapsb-r@mail.nih.gov
          This clone is available royalty-free through LNLN; contact the
          IMAGE Consortium (info@image.lnl.gov) for further information.
          Insert Length: 1426 Std Error: 0.00
          Seq primer: -40UP from Glibco
          High quality sequence stop: 465.
          Location/Qualifiers
            1..562
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:2348702"
              /lab_host="DH10B"
              /clone_lib="Soares NFL T GBC_S1"
              /note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with
              a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
              Equal amounts of plasmid DNA from three normalized
              libraries (fetal lung Nhlh19W, testis NHT, and B-cell
              NCI CGAP GCBI) were mixed, and ss circles were made in
              vitro. Following HAP purification, this DNA was used as
              tracer in a subtractive hybridization reaction. The driver
              was PCR-amplified cDNAs from pools of 5,000 clones made
              from the same 3 libraries. The pools consisted of
              I.M.A.G.E. clones 297480-302087, 682632-687239,
              726408-728711, and 729096-731399. Subtraction by Bento
              Soares and M. Fatima Bonaldo."
BASE COUNT 134 a 148 c 161 g 119 t
ORIGIN
Query Match 16.8%; Score 536.8; DB 9; Length 562;
Best Local Similarity 99.3%; Pred. No. 7.9e-150;
Matches 560; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

```

Db      444 CTTCTCGGCTCATATCATATGACAGACGCGGCTGTGTGCTGACCTGAATCAGAG 385
QY      967 CCCAAGCCAGCGAGAGAGAGCTTCCTTGCGCCCATGAGAGGTAAGCGAGCCTTGTA 1026
Db      384 CCCAAGCCAGCGAGAGAGAGCTTCCTTGCGCCCATGAGAGGTAAGCGAGCCTTGTA 325
QY      1027 TGGAAAAGTCCCTTGCGCCCTTAAGTATGTTCCGCAATGGTCCGGCTTGATCCTATCC 1086
Db      324 TGGAAAAGTCCCTTGCGCCCTTAAGTATGTTCCGCAATGGTCCGGCTTGATCCTATCC 265
QY      1087 ATGTGCCAATTTGTAAGTCTCAAGTATGAGAGAAATGTAATGACGCAAGGCCCTTC 1146
Db      264 ATGTGCCAATTTGTAAGTCTCAAGTATGAGAGAAATGTAATGACGCAAGGCCCTTC 205
QY      1147 CGCAACCCAGAAATGACACATCTCTCGGGCTGACCTACAGAGGGGCTTCTGACATG 1206
Db      204 CGCAACCCAGAAATGACACATCTCTCGGGCTGACCTACAGAGGGGCTTCTGACATG 145
QY      1207 AGCTTGCGCTGGGCCCACTCTCTAGTTCCTCATTAATAAGACAGATGCTTCTGCTTC 1266
Db      144 AGCTTGCGCTGGGCCCACTCTCTAGTTCCTCATTAATAAGACAGATGCTTCTGCTTC 85
QY      1267 TCACTGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTCCCACTTCTCATTAAT 1326
Db      84 TCACTGAGGGGCTTCTGACATGAGTCTGGCTGGCCCACTCCCACTTCTCATTAAT 25
QY      1327 AAAAGACAGTGTGCTTCTCACTTG 1350
Db      24 AAAAGACAGTGTGCTTCTCACTTG 1

RESULT 12
LOCUS   AV661170      602 bp      mRNA      linear      EST 16-JAN-2002
DEFINITION AV661170 GLC Homo sapiens cDNA clone GLCSPD10.3', mRNA sequence.
ACCESSION AV661170
VERSION   AV661170.1 GI:9882184
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS   Xu,X., Huang,J., Liu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
          Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
          Shen,K., Lu,G., Fu,G., Zhong,W., Xu,S., Gu,W., Huang,W., Zhao,X.,
          Hu,G., Gu,J., Chen,Z., and Han,Z.
          Insight into hepatocellular carcinogenesis at transcriptome level
          by comparing gene expression profiles of hepatocellular carcinoma
          with those of corresponding noncancerous liver
          Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
JOURNAL   MEDLINE
PUBMED    21625106
CONTACT:  Zeguang Han
          Chinese National Human Genome Center at Shanghai
          351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
          201203, P. R. China
          Tel: 86-21-50801919(ex.45)
          Fax: 86-21-50801922
          Email: hanzg@chgc.sh.cn
          This clone is available at CHGC in Shanghai.
          Location/Qualifiers
            1..602
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="GLCSPD10"
              /tissue_type="corresponding non cancerous liver tissue"
              /dev_stage="Adult"
              /lab_host="SOLR"
              /clone_lib="GLC"
              /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
              XhoI"
FEATURES
source

```

```

QY 837 GAGAAAGAGTGGGAGCCCTCATGAGTATCTGAGCTGTGAGTGGCCCTTATC 896
Db 489 GAAAGAGAGTGGGAGCCCTCATGAGTATCTGAGCTGTGAGTGGCCCTTATC 430
QY 897 TGGCCTGTGGTCTTCTGGCTTCATATCATGAGAGAGCGCGGTGTGTGGCTGAGCTG 956
Db 429 TGGCCTGTGGTCTTCTGGCTTCATATCATGAGAGAGCGCGGTGTGTGGCTGAGCTG 370
QY 957 AAGCAGAGAGCCGAGAGCTGAGAGAGCTTCTTGGCCCATGAGAGTGAAGCGA 1016
Db 369 AAGCAGAGAGCCGAGAGCTGAGAGAGCTTCTTGGCCCATGAGAGTGAAGCGA 310
QY 1017 GCCTGTGTATGAGAAAGTCCCTTGAGCCCTAATCATGTTCCGCAATGAGTCCGAGCTTG 1076
Db 309 GCCTGTGTATGAGAAAGTCCCTTGAGCCCTAATCATGTTCCGCAATGAGTCCGAGCTTG 250
QY 1077 TACCTCATCATGAGTCCCAATTTGTACTGTACAGTATGTGAGAAACTGAATGAGCC 1136
Db 249 TACCTCATCATGAGTCCCAATTTGTACTGTACAGTATGTGAGAAACTGAATGAGCC 190
QY 1137 AAGGCTCTTCCGCAACCCGAGAAATGACAGTCTCTGGGCTGCACTCACTGAGGGGCC 1196
Db 189 AAGGCTCTTCCGCAACCCGAGAAATGACAGTCTCTGGGCTGCACTCACTGAGGGGCC 130
QY 1197 TTCTGACATGAGTCTGAGCTGAGCCCACTCTCTAGTCTCTCATPATAAAGAGATTGCT 1256
Db 129 TTCTGACATGAGTCTGAGCTGAGCCCACTCTCTAGTCTCTCATPATAAAGAGATTGCT 70
QY 1257 TCT-TGCTTCTCATGAGGGGCTTCTGTACATGAGTCTGAGCCCTGAGCCCACTCTCCAG 1315
Db 69 TCTTGCTCTCTCATGAGGGGCTTCTGTACATGAGTCTGAGCCCTGAGCCCACTCTCCAG 10
QY 1316 TTTCTCTATA 1324
Db 9 TTTCTATA 1

```

```

RESULT 10
AL532235 1051 bp mRNA linear EST 23-MAY-2003
LOCUS AL532235 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION CS0DM004YJ20 5-PRIME, mRNA sequence.
ACCESSION AL532235
VERSION AL532235.2 GI:31070067
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1051)
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12795728.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5958.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM004DE10P1&cluster=5958.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DM004DE10P1.
Location/Qualifiers
1. 1051
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YJ20"

```

```

FEATURES
source

```

```

/issue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 234 a 281 c 303 g 233 t
ORIGIN

```

```

Query Match 16.9%; Score 537; DB 9; Length 1051;
Best Local Similarity 78.3%; Pred. No. 1,2e-149;
Matches 771; Conservative 0; Mismatches 0; Indels 214; Gaps 2;

```

```

QY 2 TCTGAGCTCAGATGAGTACTGAGAGACCCGTTGACCTGGGGTGTGAGACTT 61
Db 65 TCTGAGCTCAGATGAGTACTGAGAGACCCGTTGACCTGGGGTGTGAGACTT 124
QY 62 ATGCTGATCTGAGGCAATGGCCCACTCTCTCCGACTAGTGGCCATGGAATGTTC 121
Db 125 ATGCTGATCTGAGGCAATGGCCCACTCTCTCCGACTAGTGGCCATGGAATGTTC 184
QY 122 TGAAGCGAGACCAACCCAGACCGTGAACGCTGACATGAGCTGAGACTT 181
Db 185 TGAAGCGAGACCAACCCAGACCGTGAACGCTGACATGAGCTGAGACTT 244
QY 182 TGAATCTACCACTCTGAGTGAACATGAGTCTGTTTAAAGGGAGTTGTG 241
Db 245 TGAATCTACCACTCTGAGTGAACATGAGTCTGTTTAAAGGGAGTTGTG 304
QY 242 GAAGAGTCACAATGAGGACCGGGAGTTAATCTCAGAGATGAGAAATTTCCAGCC 301
Db 305 GAAGAGTCACAATGAGGACCGGGAGTTAATCTCAGAGATGAGAAATTTCCAGCC 354
QY 302 TGTGATGCTGATTCGTCAGTCAACAGTGTCTTCTGATCAAGGGGACAAAGT 351
Db 365 TGTGATGCTGATTCGTCAGTCAACAGTGTCTTCTGATCAAGGGGACAAAGT 424
QY 362 CTGGGATATCCCTCTGAAAAGAGAAAGATACCCAAAGTGTCTCAAGATGAT 421
Db 425 CTGGGATATCCCTCTGAAAAGAGAAAGATACCCAAAGTGTCTCAAGATGAT 484
QY 422 TCTGGAATCCCATCCCACTGATGAGCTGTGGAATGTCAACCGTGAAGATGTAAGC 481
Db 485 TCTGGAATCCCATCCCACTGATGAGCTGTGGAATGTCAACCGTGAAGATGTAAGC 544
QY 482 TGAAGCGTCTCTTCTTCCA----- 502
Db 545 TGAAGCGTCTCTTCTTCCAAGGTGACCGGAGTGTCTGGAATGAGGGAAC 604
QY 503 ----- 502
Db 605 CATGAAGAGAGCTTCTGCGCAGCTGTGGAATGCTCTCTGCGCTGAGATGCTGGG 664
QY 503 ----- 502
Db 665 CCGCTACTACTGCTTCCAGGATACCAATTCCTGCGCTTGCACCTGTCAAGGGAAGGT 724
QY 503 -----AGGCA 508
Db 725 GCCTCCAGGTACCGGGAGATGTCGAGACTACTCATGACCTGCGCTGAGAGGCA 784
QY 509 TGAACACAGAGATGAGTCTGCGCATGAGAAACATACCACTAGCCCTGAGATATGG 568
Db 785 TGAACACAGAGATGAGTCTGCGCATGAGAAACATACCACTAGCCCTGAGATATGG 844
QY 569 CTGTAGCCCAATCTAGTCTGTGCTGACATGAGCTGTGAACAACCATGAGTCCACTATGC 628
Db 845 CTGTAGCCCAATCTAGTCTGTGCTGACATGAGCTGTGAACAACCATGAGTCCACTATGC 904
QY 629 CTTCAGTGGAGCCCACTACTGAGCTGTGACACCAAGCCGGAGTGGCTGACATGCTGGCC 688

```

AUTHORS	Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z., and Han, Z.
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE	21525106
PUBMED	11752456
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@hgc.sh.cn This clone is available at CHGC in Shanghai.
FEATURES	Location/Qualifiers
source	1..704 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="GLAAH08" /tissue_type="corresponding non cancerous liver tissue" /dev_string="Adult" /lab_host="SOAR" /clone_lib="GLA" /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	167 a 181 c 197 g 155 t 4 others
ORIGIN	
Query Match	21.0%; Score 668; DB 9; Length 704;
Best Local Similarity	98.0%; Pred. No. 2.5e-189;
Matches 685; Conservative 0; Mismatches 13; Indels 1; Gaps 1;	
659	CACGAGCGGGATGGCTGCGATAGCTGGCCCTATGCTCATGAGGCCACAGGCTTTC
703	CACAGCGCGAATGGCTGGCATAGCTGGCCCTATGCTCATGAGGCCACAGGCTTTC
719	AGCAGTGATGCTGCTTTTCTTCTGGGAAGAAAATCTT-ATCTGATCAGGACCCAG
643	AGCAGTGATGCTGCTTTTCTTCTGGAAGAAAATCTTATCTGATCAGGACCCAG
778	TATATGTTTCTGACAAAGGAGGCTTATCCTAGTAAGGCTTATCCGAAGCGGCTG
583	TATATGTTTCTGACAAAGGAGGCTTATCCTAGTAAGCGGTTATCCGAAGCGGCTG
838	AGAAGGAAGTGGGACCCCTCATGAGGATATCTCTGAGCTCTGTGATGCGGCTTATCT
523	AGAAGGAAGTGGGACCCCTCATGAGGATATCTCTGAGCTCTGTGATGCGGCTTATCT
898	GCCCTGGGCTTTCTGCGCTCATATCATGAGACGCGCGCTGTGCTGACCTGA
463	GCCCTGGGCTTTCTGCGCTCATATCATGAGACGCGCGCTGTGCTGACCTGA
958	AGTACAGAGCCCAAGCCACGTGACACAGGCTTCTTGGGCCCATAGAAAGTAGACGAG
403	AGTACAGAGCCCAAGCCACGTGACACAGGCTTCTTGGGCCCATAGAAAGTAGACGAG
1018	CCTTGTGATGAAAGAGTCCCTGGCCCTTAATCAATGTTTGGCCCAATGATCCCGGCTTG
343	CCTTGTGATGAAAGAGTCCCTGGCCCTTAATCAATGTTTGGCCCAATGATCCCGGCTTG
1078	ACCTCATCATGATGATCCCAATTTGTAATGCTGTAACATGATGTGAGAAATGATGACCA
283	ACCTCATCATGATGATCCCAATTTGTAATGCTGTAACATGATGTGAGAAATGATGACCA
1138	AGGCGCTTGGCAACCCCAAGATGACCAAGTCTCTGGGCTGACCTCACTGAGGGGCTT
223	AGGCGCTTGGCAACCCCAAGATGACCAAGTCTCTGGGCTGACCTCACTGAGGGGCTT

QY	1198	TCTGACATGATGCTGGCGCTGGCCCCACCTCCTAGTCTCTCATATTAAGACAGATGCTT	125
Db	163	TCTGACATGATGCTGGCGCTGGCCCCACCTCCTAGTCTCTCATATTAAGACAGATGCTT	104
QY	1258	CTTGCGCTTCTCACTGAGGGGCTTCTGACATAGTCTGGCGCCCACTCCCACTT	131
Db	103	CTTGCGCTTCTCACTGAGGGGCTTCTGACATAGTCTGGCGCCCACTCCCACTT	44
QY	1318	TCTCATATTAAGACAGATGCTGCTTCTTCACTTGAATCAA	1356
Db	43	TCTCATATTAAGACAGATGCTGCTTCTTCACTTGAATCAA	5
RESULT 9	877 bp	linear	EST 21-APR-2001
LOCUS	BG196922		
DEFINITION	RS116153 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BG196922		
VERSION	BG196922.1	GI:13718609	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McEllisot,K., Booser,S., Mays,R., Smith,E., Veloso,N., Klinka,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Darrig,J. and Ducar,M.		
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression		
JOURNAL	Nat. Biotechnol. 19 (5), 440-445 (2001)		
MEDLINE	21227151		
PUBMED	11329013		
COMMENT	Contact: Scott J. Cain Atherys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scain@atherys.com		
FEATURES	High quality sequence stop. 326.		
SOURCE	Location/Qualifiers 1..877 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /cell_line="HT1080" /clone_id="Atherys RAGE Library" /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."		
BASE COUNT	206 a 213 c 289 g 169 t		
ORIGIN			
Query Match	17.2%; Score 549; DB 10; Length 877;		
Best Local Similarity	90.7%; Pred. No. 2,46-153;		
Matches 607; Conservative 0; Mismatches 60; Indels 2; Gaps 2;			
QY	658	ACACGAGCGGAGAGGCTGGCATAGTGGCCCATTTGCTCATGAGGCCAGGGATCCTT	717
Db	669	ACCAACCTTGGAGAGGCTTGAATAGAGCACCAATGTTTCATCAGGGGCCAGTCCCT	610
QY	718	CAGCAGTGAATGCTGCTTTCTTCTGGAGAG-AAAACCTATCTGTTCAGGAGCCAG	776
Db	609	AGCAGTGAAGTCTCCTCTTCTGGAGACATATACTTATTTGTTCAAGGGCCCCAG	550
QY	777	GTAATATGCTTCTGACAAAGAGGCTATACCTTAAGCGGTTATCCGAAAGCGCTG	836
Db	549	GTATATTTTCTGCTCAAGGAGGCTATCCCTTATACCGGTATCTCAAGCGGCTG	490

```

Db      451 ATGGATATATCTGAGTCTGTGGATGCGGCTTATCTGCCCTGCGGCTTCTCGGCTCC 392
Qy      919 ATATCATGAGAGACGGCGGCTGTGTGGCTGAGACCTGAAGTACAGAGACCCCAACCACT 978
Db      391 ATATCATGAGAGACGGCGGCTGTGTGGCTGAGACCTGAAGTACAGAGACCCCAACCACT 332
Qy      979 GACAGAGCTTCTTGGCGCCCATGAGAGGTAAGCGAGGCTTGTATGTAAGTAAGTCCC 1038
Db      331 GACAGAGCTTCTTGGCGCCCATGAGAGGTAAGCGAGGCTTGTATGTAAGTAAGTCCC 272
Qy      1039 TTGGCCCTTAATCTCATGTTCCGCAATGATGTCGCCCTTGTACCTTCACTGATGTCCTCAAT 1098
Db      271 TTGGCCCTTAATCTCATGTTCCGCAATGATGTCGCCCTTGTACCTTCACTGATGTCCTCAAT 212
Qy      1099 TGTACTGCTACAGTATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1158
Db      211 TGTACTGCTACAGTATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 152
Qy      1159 ATGTGACCACTCTCTGAGGCTGACACTGAGAGGAGGCTTCTGACATGAGTCTGAGCTTGG 1218
Db      151 ATGTGACCACTCTCTGAGGCTGACACTGAGAGGAGGCTTCTGACATGAGTCTGAGCTTGG 92
Qy      1219 CCCCACTCTCTGAGTCTCTCATATATTAAGACAGATTTGCTTCTGCTTCTGACATGAGGAGC 1278
Db      91 CCCCACTCTCTGAGTCTCTCATATTAAGACAGATTTGCTTCTGCTTCTGACATGAGGAGC 32
Qy      1279 CTCTGACATGAGTCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1307
Db      31 CTCTGACATGAGTATATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 3

```

```

RESULT 7
LOCUS   BX448764/c 1156 bp mRNA linear EST 22-MAY-2003
DEFINITION
BX448764 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
ACCESSION
BX448764
VERSION
BX448764.1 GI:31021961
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1. (bases 1 to 1156)
Liang, L., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5958.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AM011CA12NP1&cluster=5958.r. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AM011CA12NP1.
Location/Qualifiers
1. 1156
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM011YB23"
/rna_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: Liver; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6

```

## FEATURES

SOURCE

```

1. 1156
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM011YB23"
/rna_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: Liver; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6

```

```

BASE COUNT      vector. Library was not normalized."
ORIGIN          264 a 260 c 301 g 240 t 91 others

```

```

Query Match      21.6%; Score 688.8; DB 13; Length 1156;
Best Local Similarity 97.3%; Pred. No. 2e-195;
Matches 745; Conservative 4; Mismatches 12; Indels 5; Gaps 5;

```

```

Qy      503 AGGCGCATGAGACAGAGATGAGGAGCTGGCCATGAGGAGACAGTACCCATGAGCCCTGAGTA 562
Db      765 AGGCGCATGAGACAGAGATGAGGAGCTGGCCATGAGGAGACAGTACCCATGAGCCCTGAGTA 706
Qy      563 TATGCGCTGAGCCCACTATGTTCTTGTCTGCACTGAGCTGACACCAATGAGTCCAC 622
Db      705 TATGCGCTGAGCCCACTATGTTCTTGTCTGAC - KAGGTGACACACAT - GTGCCAC 648
Qy      623 CTATGCTTCAAGTGGAGCCCACTATGAGGCTTGGACACCAAGCGGAGATGAGTGGCATAG 662
Db      647 CTATGCTTCAAGT - GAGCCCACTATGAGGCTTGGACACCAAGCGGAGATGAGTGGCATAG 589
Qy      683 CTGGCCCATTTGCTCATCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 742
Db      588 CTGGCCCATTTGCTCATCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 530
Qy      743 GGAAGAAAACTCTATCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 802
Db      529 GGAAGAAAACTCTATCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 470
Qy      803 CTATACCTTATGTAAGGAGGATATCCGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862
Db      469 CTATACCTTATGTAAGGAGGATATCCGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
Qy      863 GATTATCCGAGACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922
Db      409 GATTATCCGAGACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
Qy      923 CATGAGAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982
Db      349 CATGAGAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290
Qy      983 AGAGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1042
Db      289 AGAGCTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230
Qy      1043 CCTTAATCAATGTTCCGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1102
Db      229 CCTTAATCAATGTTCCGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 170
Qy      1103 CTGCTACAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1162
Db      169 CTGCTACAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 110
Qy      1163 GACCACTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1222
Db      109 GACCACTTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 50
Qy      1223 ACTCTAGTTC - CTATATATAAGACAGATTTCTTCTGCTTCT 1267
Db      49 ACTCTAGTTCCTCATATATAAGACAGATTTCTTCTGCTTCT 4

```

## RESULT 8

AV645403/c

```

LOCUS   AV645403 704 bp mRNA linear EST 15-JAN-2002
DEFINITION
AV645403 GLA Homo sapiens cDNA clone GLAABH08.3, mRNA sequence.
ACCESSION
AV645403
VERSION
AV645403.1 GI:9866417
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 704)

```



/tissue\_type="FETAL BRAIN"  
/dev\_stage="Fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 243 a 306 c 340 g 268 t 44 others  
ORIGIN

Query Match 23.7%; Score 756.6; DB 13; Length 1201;  
Best Local Similarity 94.9%; Pred. No. 7.4e-216;  
Matches 787; Conservative 13; Mismatches 27; Indels 2; Gaps 2;

QY 503 AGGCGATGACACAGAAATGGAGCTGGCCATGGGAAAGTACCCACCATGGCCCTGAGTA 562  
DB 272 AGGCGATGACACAGAAATGGAGCTGGCCATGGGAAAGTACCCACCATGGCCCTGAGTA 331  
QY 563 TATGCGCTGAGGCCACATGATGCTGTCGACCTGACCTGACCAACCATGGTGGCAC 622  
DB 332 TATGCGCTGAGGCCACATGATGCTGTCGACCTGACCTGACCAACCATGGTGGCAC 391  
QY 623 CTATGCTTCACTGAGGACCACTACTGCGCTGACACCAAGCGGAGTGGCTGACATAG 682  
DB 392 CTATGCTTCACTGAGGACCACTACTGCGCTGACACCAAGCGGAGTGGCTGACATAG 451  
QY 683 CTGCGCCATGCTCATGATGAGGCGCCAGGGCTCTTCAAGAGTGGATGCTCCCTTTCTG 742  
DB 452 CTGCGCCATGCTCATGATGAGGCGCCAGGGCTCTTCAAGAGTGGATGCTCCCTTTCTG 511  
QY 743 GGAAGAAAACTCTATGCTGTCAGAGGACCAAGGTATGCTCTCTGCAAAAGGAGG 802  
DB 512 GGAAGAAAACTCTATGCTGTCAGAGGACCAAGGTATGCTCTCTGCAAAAGGAGG 571  
QY 803 CTATACCTTATAGCGGTTATCCGAAAGCGGCTGAGAAAGAAAGTCCGACCCCTCATG 862  
DB 572 CTATACCTTATAGCGGTTATCCGAAAGCGGCTGAGAAAGAAAGTCCGACCCCTCATG 631  
QY 863 GATTATCTGGAATCTGTGAGATGGGCTTTATCTGCTGGCTGCTTCTGCTGCTCATAT 922  
DB 632 GATTATCTGGAATCTGTGAGATGGGCTTTATCTGCTGGCTGCTTCTGCTGCTCATAT 691  
QY 923 CATGGACGAGCGGCGCTGTGCTGAGACCTGAATCGAGAGCCCAAGCCACGTGGAC 982  
DB 692 CATGGACGAGCGGCGCTGTGCTGAGACCTGAATCGAGAGCCCAAGCCACGTGGAC 751  
QY 983 AGAGCTTCTTGGCCCATGAGAAAGTGAAGAGAGCTTGTGATGAGAAAGTCCCTTGG 1042  
DB 752 AGAGCTTCTTGGCCCATGAGAAAGTGAAGAGAGCTTGTGATGAGAAAGTCCCTTGG 811  
QY 1043 CCTTAATCATGTTCCGCAATGATCCGCTTGTATCTCATTCATGCTCCCAATTTGTA 1102  
DB 812 CCTTAATCATGTTCCGCAATGATCCGCTTGTATCTCATTCATGCTCCCAATTTGTA 871  
QY 1103 CTGCTACAGATGATGAGAACTGAATGAGCCCAAGCCCTCCGAAACCCCAAGATG 1162  
DB 872 CTGCTACAGATGATGAGAACTGAATGAGCCCAAGCCCTCCGAAACCCCAAGATG 931  
QY 1163 GACCAAGCTCTGGGCTGACACTGAGAGGCGCTTCTGACATGATGATGGGCTGGCC 1221  
DB 932 GACCAAGCTCTGGGCTGACACTGAGAGGCGCTTCTGACATGATGATGGGCTGGCC 991  
QY 1222 GACC-TCCTAGTTCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGGGCT 1280  
DB 992 GACC-TCCTAGTTCTCATATAAAGACAGATTGCTTCTGCTTCACTGAGGGGCT 1051  
QY 1281 TCTGACATGATGCTGGGCTGGCCCACTCCCAAGTTTCTCATATAA 1329  
DB 1052 TCTTGAATAGTGTGSCCTGGCCCACTCCCGMGTATTATATAAARMAA 1100

RESULT 6

BX405876/c  
LOCUS 1201 bp mRNA linear EST 13-MAY-2003  
DEFINITION BX405876 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
ACCESSION CS0DM009Y021.3-PRIME, mRNA sequence.  
VERSION BX405876  
KEYWORDS BX405876.1 GI:30652298  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r for  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DM009AH11NP1&cluster=5958.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID: CS0DM009AH11NP1.

FEATURES  
source  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM009Y021"  
/issue\_type="FETAL LIVER"  
/dev\_stage="Fetal"  
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 245 a 303 c 303 g 265 t 120 others  
ORIGIN

Query Match 23.3%; Score 743.4; DB 13; Length 1201;  
Best Local Similarity 94.9%; Pred. No. 7.1e-212;  
Matches 768; Conservative 18; Mismatches 19; Indels 4; Gaps 2;

QY 503 AGGCGATGACACAGAAATGGAGCTGGCCATGGGAAAGTACCCACCATGGCCCTGAGTA 562  
DB 811 AGGCGATGACACAGAAATGGAGCTGGCCATGGGAAAGTACCCACCATGGCCCTGAGTA 752  
QY 563 TATGCGCTGAGGCCACATGATGCTGTCGACCTGACCTGACCAACCATGGTGGCAC 622  
DB 752 TATGCGCTGAGGCCACATGATGCTGTCGACCTGACCTGACCAACCATGGTGGCAC 692  
QY 623 CTATGCTTCACTGAGGACCACTACTGCGCTGACACCAAGCGGAGTGGCTGACATAG 682  
DB 692 CTATGCTTCACTGAGGACCACTACTGCGCTGACACCAAGCGGAGTGGCTGACATAG 632  
QY 683 CTGCGCCATGCTCATGATGAGGCGCCAGGGCTCTTCAAGAGTGGATGCTGCTTCTG 742  
DB 632 CTGCGCCATGCTCATGATGAGGCGCCAGGGCTCTTCAAGAGTGGATGCTGCTTCTG 572  
QY 743 GGAAGAAAACTCTATGCTGTCAGAGGACCAAGGTATGCTCTCTGCAAAAGG 799  
DB 572 GGAAGAAAACTCTATGCTGTCAGAGGACCAAGGTATGCTCTCTGCAAAAGG 512  
QY 800 -AGGCTATACCTTATGAGGCTTATCCGAAAGGCTGAGAAAGAAAGTGGGACCCCTC 858  
DB 511 TAGGCTATACCTTATGAGGCTTATCCGAAAGGCTGAGAAAGAAAGTGGGACCCCTC 452  
QY 859 ATGGATTAATCCGAGCTGTGAGATGGGCGCTTATATGCTGCGGCTGCTC 918

QY 1283 TGACATGAGTCTGCGCCGCGCCACCGCC-CGAGTTTCATATAATAAGACAGATTGCT 1340  
 Db 66 TGACATGAGTCTGCGCCGCGCCACCGCC-CGAGTTTCATATAATAAGACAGATTGCT 8

RESULT 4  
 AL564746/c  
 LOCUS AL564746  
 DEFINITION AL564746 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
 CS0DM007YC08 3-PRIME, mRNA sequence.  
 ACCESSION AL564746  
 VERSION AL564746.2 GI:31288721  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1201)  
 Full-length cDNA libraries and normalization  
 Unpublished  
 On Feb 16, 2001 this sequence version replaced gi:12915461.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 5958.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DM007B04NP1&cluster=5958.r. Contact :  
 Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DM007B04NP1.  
 Location/Qualifiers  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DM007YC08"  
 /issue\_type="FETAL LIVER"  
 /dev\_stage="fetal"  
 /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 vector. Library was not normalized."  
 BASE COUNT 278 a 291 c 323 g 267 t 42 others  
 ORIGIN

Query Match 25.2%; Score 802.6; DB 9; Length 1201;  
 Best Local Similarity 98.8%; Pred. No. 9.8e-230;  
 Matches 825; Conservative 3; Mismatches 5; Indels 2; Gaps 2;

QY 503 AGGCCATGACACAGAGATGGAGCTGGCCATGGGAAACAGTACCACCATGGCCCTGAGTA 562  
 Db 836 AGGCCATGACACAGAGATGGAGCTGGCCATGGGAAACAGTACCACCATGGCCCTGAGTA 777

QY 563 TATGCGCTGAGCCACATCTAGCTTGTGTGCACTGAGAGCTGACAAACATGAGTGCAC 652  
 Db 776 TATGCGCTGAGCCACATCTAGCTTGTGTGCACTGAGAGCTGACAAACATGAGTGCAC 717

QY 623 CTATGCTTCACTGAGGACCCACTAGTGGCTGTGACACACGCGGATGGCTGAGTAG 682  
 Db 716 CTATGCTTCACTGAGGACCCACTAGTGGCTGTGACACACGCGGATGGCTGAGTAG 657

QY 683 CTGCGCCATGCTCATCTAGTGGCCCGGAGGCTCTTCAAGCAGTGAAGTCTCTTTTCTG 742  
 Db 656 CTGCGCCATGCTCATCTAGTGGCCCGGAGGCTCTTCAAGCAGTGAAGTCTCTTTTCTG 597

QY 743 GGAAGAAAACTATCTGCTCAAGGACCCAGGTATATGCTTCTTCAAGGAGGAG 802

Db 596 GGAAGAAAACTATCTGCTCAAGGACCCAGGTATATGCTTCTTCAAGGAGGAG 537  
 QY 803 CTATACCTTCACTGAGGAGGCTTATCCGAGGAGGCTGAGAAAGATCGGAGACCCCTCATG 862  
 Db 536 CTATACCTTCACTGAGGAGGCTTATCCGAGGAGGCTGAGAAAGATCGGAGACCCCTCATG 477

QY 863 GATTATCCCTGAGCTGTGAGGAGGCTTATCTGAGGAGGCTTCTCGGCTCATAT 922  
 Db 476 GATTATCCCTGAGCTGTGAGGAGGCTTATCTGAGGAGGCTTCTCGGCTCATAT 417

QY 923 CATGCGAGACGCGCGCTGTGAGCTGACCTGAGAGGAGCCCAAGCCACGAGTGA 982  
 Db 416 CATGCGAGACGCGCGCTGTGAGCTGACCTGAGAGGAGCCCAAGCCACGAGTGA 357

QY 983 AAGACTTCTTGGCCCATGAGAGGAGTGAAGGAGCTTGTATGAGAAAGTCCCTTGG 1042  
 Db 356 AAGACTTCTTGGCCCATGAGAGGAGTGAAGGAGCTTGTATGAGAAAGTCCCTTGG 297

QY 1043 CCTTACCTGAGTGTGCGCAATGAGTCCGAGCTTGTACCTCATGCAATGATCCCAATTGTA 1102  
 Db 296 CCTTACCTGAGTGTGCGCAATGAGTCCGAGCTTGTACCTCATGCAATGATCCCAATTGTA 237

QY 1103 CTGCTACAGTGAATGTGAGAAAGTGAATGACGCAAGGCTTCCGCAACCCAGATGT 1162  
 Db 236 CTGCTACAGTGAATGTGAGAAAGTGAATGACGCAAGGCTTCCGCAACCCAGATGT 177

QY 1163 GACCACTTCTGCGGCTGCACTCACTGAGGAGCTTGTGACATGAGTCTGCGCTGCCCC 1222  
 Db 176 GACCACTTCTGCGGCTGCACTCACTGAGGAGCTTGTGACATGAGTCTGCGCTGCCCC 117

QY 1223 ACCTCTATCTCTCATATAATAAGACAGATTGCTTGTGCTTCTGCAAGGAGGCTTC 1282  
 Db 116 ACCTCTATCTCTCATATAATAAGACAGATTGCTTGTGCTTCTGCAAGGAGGCTTC 57

QY 1283 TGACATGAGTCTGCGCCGCGCCACCGCC-CGAGTTTCATATAATAAGACAGATTGCT 1337  
 Db 56 TGACATGAGTCTGCGCCGCGCCACCGCC-CGAGTTTCATATAATAAGACAGATTGCT 4

RESULT 5  
 BX440692  
 LOCUS BX440692 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 CS0DF010YK01 5-PRIME, mRNA sequence.  
 ACCESSION BX440692  
 VERSION BX440692.1 GI:30781865  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1201)  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 5958.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DF010AF010P1&cluster=5958.r. Contact :  
 Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DF010AF010P1.  
 Location/Qualifiers  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF010YK01"

```

Db      607 TTTCCTGGGAGAGAACTATCTGATCGAGGACCCAGATATATGTCTTCTGACAA 548
Qy      796 AGGAGAGCTATACCTAGTAAGGGTTATCCGAAAGCGGCTGGAAGAGAGTGGGACCC 855
Db      547 AGGAGAGCTATACCTAGTAAGGGTTATCCGAAAGCGGCTGGAAGAGAGTGGGACCC 488
Qy      856 CTGATGAGATATCTGAGACTGTGAGATCGGCTTTATCTGCGCTTCTTCTGCG 915
Db      487 CTGATGAGATATCTGAGACTGTGAGATCGGCTTTATCTGCGCTTCTTCTGCG 428
Qy      916 TCATATCATGAGAGAGCGGCGCTGTGTGCTGAGACTGAAGTCAAGAGCCCAAGCCA 975
Db      427 TCATATCATGAGAGAGCGGCGCTGTGTGCTGAGACTGAAGTCAAGAGCCCAAGCCA 368
Qy      976 CGTGAAGAGCTTCTTGGCCCCCAGAGAGAGTGAAGAGCGCTTGTATGAGAAAGT 1035
Db      367 CGTGAAGAGCTTCTTGGCCCCCAGAGAGAGTGAAGAGCGCTTGTATGAGAAAGT 308
Qy      1036 CCTTGGCCCTTACTCATGTTTCCGCCAATGTCGCCGCTTGTACTCATGATGTCCTCA 1095
Db      307 CCTTGGCCCTTACTCATGTTTCCGCCAATGTCGCCGCTTGTACTCATGATGTCCTCA 248
Qy      1096 ATTTGACTGCTACAGTATGTGAGAGAACTGAATGAGCCAAAGCCCTTCCGCAACCC 1155
Db      247 ATTTGACTGCTACAGTATGTGAGAGAACTGAATGAGCCAAAGCCCTTCCGCAACCC 188
Qy      1156 AGAATGAGCAAGTCTCTGGGCTGCACTACAGAGGGGCTTCTGACATGATGTCGCC 1215
Db      187 AGAATGAGCAAGTCTCTGGGCTGCACTACAGAGGGGCTTCTGACATGATGTCGCC 128
Qy      1216 TGGCCCCACCTCTTACTTCTCATTAATAAGACAGATTGCTTCTGCTTCTGACGAG 1275
Db      127 TGGCCCCACCTCTTACTTCTCATTAATAAGACAGATTGCTTCTGCTTCTGACGAG 68
Qy      1276 GGGCTTCTGACATGATGTCGCCCTGCGCCCACTCCCACTTCTCATTAATAAGACAG 1335
Db      67 GGGCTTCTGACATGATGTCGCCCTGCGCCCACTCCCACTTCTCATTAATAAGACAG 9
Qy      1336 TTGCTTCT 1343
Db      8 TTGCTTCT 1

```

RESULT 3  
AL564329/c 1053 bp mRNA linear EST 31-MAY-2003  
LOCUS AL564329 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
DEFINITION CS0DM004YJ20 3-PRIME, mRNA sequence.  
ACCESSION AL564329  
VERSION AL564329.2 GI:31288307  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1053)  
Li,W.B., Gruber,C., Jessee,J. and Polyes,D.  
Full-length cDNA libraries and normalization.  
Unpublished  
On Feb 16, 2001 this sequence version replaced gi:12914625.  
Contact: Genoscope  
Genoscope - Centre National de Sequenage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgl-bin/cluster.cgi?seq=CS0DM004DE10NP1&cluster=5958.r. Contact :  
Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/InvitrogenCorporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DM004DE10NP1.

```

FEATURES
source
Location/Qualifiers
1..1053
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YJ20"
/issue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_1ib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT
237 a 286 c 306 g 210 t 14 others
ORIGIN
Query Match 25.6%; Score 816; DB 9; Length 1053;
Best Local Similarity 98.6%; Pred. No. 8e-234;
Matches 827; Conservative 3; Mismatches 8; Indels 1; Gaps 1;
Qy 503 AGGCAATGACACAGAAATGGGACTGGCCATGGGAAACAGTACCACATGGCCCTGAGTA 562
Db 846 AGGCCATGACACAGAAATGGGACTGGCCATGGGAAACAGTACCACATGGCCCTGAGTA 787
Qy 563 TATGGCTGAGCCCATCTAGTCTTGTCTGACTGAGAGTCTGACACATGATGTCGAC 622
Db 786 TATGGCTGAGCCCATCTAGTCTTGTCTGACTGAGAGTCTGACACATGATGTCGAC 727
Qy 623 CTATGCTTCACTGAGGACCACTACTGAGGCTGACACACAGCCGGAGTGGCTGACATG 682
Db 726 CTATGCTTCACTGAGGACCACTACTGAGGCTGACACACAGCCGGAGTGGCTGACATG 667
Qy 683 CTGAGCCATGCTCATCATGAGGCCCGAGGCTCTTACAGAGTGAATGTCGCTTTCTCG 742
Db 666 CTGAGCCATGCTCATCATGAGGCCCGAGGCTCTTACAGAGTGAATGTCGCTTTCTCG 607
Qy 743 GGAAGAAAACCTATCTGCTCAAGGACCCAGGATATGCTTCTGACAAAGGAGG 802
Db 606 GGAAGAAAACCTATCTGCTCAAGGACCCAGGATATGCTTCTGACAAAGGAGG 547
Qy 803 CTATACCTAGTAAGCGATTATCCGAAGCGGCTGAGAGAGTGGGACCCCTCATG 862
Db 546 CTATACCTAGTAAGCGATTATCCGAAGCGGCTGAGAGAGTGGGACCCCTCATG 487
Qy 863 GATTATCTGAGACTCTGTGAGATGCGGCTTTATCTGCGCTTCTTCTGCTCATAT 922
Db 486 GATTATCTGAGACTCTGTGAGATGCGGCTTTATCTGCGCTTCTTCTGCTCATAT 427
Qy 923 CATGAGAGAGCGGCGCTGTGAGTGGCTGACCTGAAGTCAAGAGCCCAAGCGTGAAC 982
Db 426 CATGAGAGAGCGGCGCTGTGAGTGGCTGACCTGAAGTCAAGAGCCCAAGCGTGAAC 367
Qy 983 AGAGCTTCTTGGCCCATGAGAGAGTGAACGAGGCTTGTATGAGAAAGTCCCTTGG 1042
Db 366 AGAGCTTCTTGGCCCATGAGAGAGTGAACGAGGCTTGTATGAGAAAGTCCCTTGG 307
Qy 1043 CCTTAACTGATGTTCCGCAATGTCGCCGCTTGTACTCATCATGATGTCCTCAATTGTA 1102
Db 306 CCTTAACTGATGTTCCGCAATGTCGCCGCTTGTACTCATCATGATGTCCTCAATTGTA 247
Qy 1103 CTGCTACAGTATGAGAGAACTGAATGACAGCCCAAGGCTTCCGCAACCCAGAAATG 1162
Db 246 CTGCTACAGTATGAGAGAACTGAATGACAGCCCAAGGCTTCCGCAACCCAGAAATG 187
Qy 1163 GACCACTCTCTGAGGCTGACTCACTGAGGAGGCTTCTGACATGAGTCTGAGGCTGCCC 1222
Db 186 GACCACTCTCTGAGGCTGACTCACTGAGGAGGCTTCTGACATGAGTCTGAGGCTGCCC 127
Qy 1223 ACTCTTACTTCTCATTAATAAGACAGATTGCTTCTTCTGCTTCTCATGAGGAGGCTTTC 1282
Db 126 ACTCTTACTTCTCATTAATAAGACAGATTGCTTCTTCTGCTTCTCATGAGGAGGCTTTC 67

```

Faraday Avenue Genoscope sequence ID : CS0AJO10BC12NP1.

# FEATURES

source

1. 992  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ010YB24"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/clone\_1b="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 228 a 270 c 293 g 199 t 2 others

ORIGIN

Query Match 26.2%; Score 834.2; DB 13; Length 992;  
Best Local Similarity 99.8%; Pred. No. 2.4e-239;  
Matches 833; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

503 AGGCCATGACACAGAGATGGAATGAGCTGGCCATGGGAAACAGTACCACCATGGCCCTGAGTA 562  
835 AGGCCATGACACAGAGATGGAATGAGCTGGCCATGGGAAACAGTACCACCATGGCCCTGAGTA 776  
563 TATGCGCTGATGACCATCTTCTTGTCTGACATGACCTCTGACACCATGGTGGCAC 622  
775 TATGCGCTGATGACCATCTTCTTGTCTGACATGACCTCTGACACCATGGTGGCAC 716  
623 CTATGCGCTGATGACCATCTTCTTGTCTGACATGACCTCTGACACCATGGTGGCAC 682  
715 CTATGCGCTGATGACCATCTTCTTGTCTGACATGACCTCTGACACCATGGTGGCAC 656  
683 CTGCGCCATGTCATCATGAGTGGCCCGAGGCTCTTCAAGAGTGGTCTCTTCTG 742  
655 CTGCGCCATGTCATCATGAGTGGCCCGAGGCTCTTCAAGAGTGGTCTCTTCTG 596  
743 GGAAGAAAACTCTATCTGCTGACAGGCGACCCAGTATATGCTTCTGACAAAGGAGG 802  
595 GGAAGAAAACTCTATCTGCTGACAGGCGACCCAGTATATGCTTCTGACAAAGGAGG 536  
803 CTATACCCCTAGTAAAGGCTTACGAAAGGCGCTGAGAAAGAGTGGGAGCCCTCATG 862  
535 CTATACCCCTAGTAAAGGCTTACGAAAGGCGCTGAGAAAGAGTGGGAGCCCTCATG 476  
863 GATTATCTGAGCTCTGTGATGCGGCTTATCTGCGCTGCTTCTGCGCTCATAT 922  
475 GATTATCTGAGCTCTGTGATGCGGCTTATCTGCGCTGCTTCTGCGCTCATAT 416  
923 CATGGACAGCGCGCTGTGTGCTGACCTGAACTGAGAGCCCAAGCCACGTGGAC 982  
415 CATGGACAGCGCGCTGTGTGCTGACCTGAACTGAGAGCCCAAGCCACGTGGAC 356  
983 AGAGCTTCTGCGCCCATAGAAAGGAGAGCGAGCTGTGATGAGAAAGTCCCTTG 1042  
355 AGAGCTTCTGCGCCCATAGAAAGGAGAGCGAGCTGTGATGAGAAAGTCCCTTG 296  
1043 CCTTAATCTATGTTCCGCAATGATGCTCCGCTTGTACTCATCATGATGCTCCATTTGTA 1102  
295 CCTTAATCTATGTTCCGCAATGATGCTCCGCTTGTACTCATCATGATGCTCCATTTGTA 236  
1103 CTCTCACTGATGATGAGAACTGAATGACCAAGCCCTTCCGCAACCCCAAGATGT 1162  
235 CTCTCACTGATGATGAGAACTGAATGACCAAGCCCTTCCGCAACCCCAAGATGT 176  
1163 GACCAAGCTCTGCGGCTGACATCACTGAGGAGGCTTCTGACATGAGTCTGGCTGGCCC 1222  
175 GACCAAGCTCTGCGGCTGACATCACTGAGGAGGCTTCTGACATGAGTCTGGCTGGCCC 116  
1223 ACCTCTAGTCTCTCATATATGAACAGATGCTTCTTCTGCTCTCACTGAGGAGGCTTCT 1282  
115 ACCTCTAGTCTCTCATATATGAACAGATGCTTCTTCTGCTCTCACTGAGGAGGCTTCT 56

QY 1283 TGACATGAGCTGCGCTGGCCGCCACCTGCCAGTTTCTCATATATGAACAGATTT 1337  
DB 55 TGACATGAGCTGCGCTGGCCGCCACCTGCCAGTTTCTCATATATGAACAGATTT 1

## RESULT 2

AL534854/c 1201 bp mRNA 1linear EST 12-MAY-2003  
LOCUS AL534854 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
DEFINITION CS0DF006Y122 3-PRIME, mRNA sequence.  
ACCESSION AL534854  
VERSION AL534854.2 GI:30541057  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessup, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12798347.  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5958.r For more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DF006Y122&cluster=5958.r. Contact :  
Feng Liang Email: fliang@life.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF006BE1NP1.

## FEATURES

source

1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF006Y122"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_1b="Homo sapiens FETAL BRAIN"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 295 a 266 c 359 g 231 t 50 others

Query Match 25.8%; Score 822.4; DB 9; Length 1201;  
Best Local Similarity 98.5%; Pred. No. 1e-235;  
Matches 835; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

496 TCTTCCAGGCCATGACACAGAGATGGAATGAGCTGGCCATGGGAAACAGTACCACCATGGCC 555  
847 TCTTCCAGGCCATGACACAGAGATGGAATGAGCTGGCCATGGGAAACAGTACCACCATGGCC 788  
556 CTGAGTATATGCGCTGATGACCATCTAGTCTGTGCTGACATGAGTCTGACAAACATG 615  
787 CTGAGTATATGCGCTGATGACCATCTAGTCTGTGCTGACATGAGTCTGACAAACATG 728  
616 GTGCCACTATGCTCTTCACTGAGGAGCCCACTACTGCGCTGTGACACCAAGCCGAGTGGCT 675  
727 GTGCCACTATGCTCTTCACTGAGGAGCCCACTACTGCGCTGTGACACCAAGCCGAGTGGCT 668  
676 GCGATATGCTGGCCCATATGCTCATCACTGAGGAGCCCAAGGCTCTTCAAGAGTATGCTGCT 735  
667 GCGATATGCTGGCCCATATGCTCATCACTGAGGAGCCCAAGGCTCTTCAAGAGTATGCTGCT 608  
736 TTTCTGAGGAGAAAACTCTATCTGCTCAAGGAGCCCAAGGATATATGCTTCTGACAA 795

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 10:53:38 ; Search time 3986.5 Seconds  
(without alignments)  
19424.068 Million cell updates/sec

Title: US-09-900-448-1

Perfect score: 3186  
Sequence: 1 cctcgagctcagcatgct.....accctttcgtgtcttcca 3186

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_rtd:\*  
26: em\_ges\_pbg:\*  
27: em\_ges\_vrt1:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	834.2	26.2	992	13	BX325177 BX325177
C 2	822.4	25.8	1201	9	AL534854 AL534854
C 3	816	25.6	1053	9	AL564329 AL564329
C 4	802.6	25.2	1201	9	AL564746 AL564746

5	756.6	23.7	1201	13	BX440692	BX440692
C 6	743.4	23.3	1201	13	BX405876	BX405876
C 7	688.8	21.6	1156	13	BX448764	BX448764
C 8	668	21.0	704	9	AV645403	AV645403
C 9	549	17.2	877	10	BG196922	BG196922
C 10	537	16.9	1051	9	AL532235	AL532235
C 11	536.8	16.7	562	9	AI798878	AI798878
C 12	533.2	16.8	602	9	AV661170	AV661170
C 13	531.2	16.7	914	13	BX412777	BX412777
C 14	530.6	16.7	631	9	AV653336	AV653336
C 15	529.6	16.6	548	9	AA705094	AA705094
C 16	527.8	16.6	551	14	CB154152	CB154152
C 17	506.4	15.9	582	9	AI133162	AI133162
C 18	505.2	15.9	542	2	HSN087830	HSN087830
C 19	505.2	15.9	542	14	CB111971	CB111971
C 20	505.2	15.9	546	14	CB120347	CB120347
C 21	504.8	15.8	1201	13	BX448765	BX448765
C 22	504.2	15.8	1061	12	BM921160	BM921160
C 23	503.6	15.8	540	14	CB113371	CB113371
C 24	503.6	15.8	657	14	CB141709	CB141709
C 25	502.2	15.8	989	13	BX325178	BX325178
C 26	500.8	15.7	569	9	AV661177	AV661177
C 27	500.4	15.7	548	14	CB147228	CB147228
C 28	499.8	15.7	1201	13	BX463236	BX463236
C 29	494.2	15.5	693	10	BG617908	BG617908
C 30	484	15.5	630	10	AW965055	AW965055
C 31	491.6	15.4	791	10	BG617459	BG617459
C 32	489.4	15.4	570	10	BG616601	BG616601
C 33	486.6	15.3	657	12	BM780523	BM780523
C 34	486.2	15.3	545	14	TS4902	TS4902
C 35	482.8	15.2	1201	9	AL564442	AL564442
C 36	481.4	15.1	498	14	CB147209	CB147209
C 37	479.8	15.1	488	14	CB154135	CB154135
C 38	478.8	15.0	482	14	CB154307	CB154307
C 39	465.8	14.6	485	9	AA514026	AA514026
C 40	460.6	14.5	1201	13	BX405877	BX405877
C 41	459	14.4	784	14	CB165194	CB165194
C 42	457.6	14.4	503	14	T71372	T71372
C 43	456	14.3	749	9	AV693145	AV693145
C 44	456	14.3	890	10	BF126689	BF126689
C 45	454.6	14.3	620	12	BM780657	BM780657

#### ALIGNMENTS

RESULT 1  
BX325177/c  
LOCUS BX325177 Homo sapiens T CELLS (JURKAT CELL LINE) COI 10-NORMALIZED  
DEFINITION Homo sapiens CDNA clone CSODJ010YK24.3-PRIME, mRNA sequence.  
ACCESSION BX325177  
VERSION BX325177.1 GI:30340447  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 992)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5958.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODJ010BC12NP1c1cluster=5958.r. Contact :  
Peng Liang Email : fliang@lifeitech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600



Matches 468; Conservative 0; Mismatches 29; Indels 15; Gaps 13;

QY 839 GAAGAGAGTGGAGACCCCTGATGAGATTCCTGAGACTGTGAGAGCGGCTTTATCTG 898  
DB 498 GAAAGAGAGTGGAGACCCCTGATGAGATTCCTGAGACTGTGAGAGCGGCTTTATCTG 440  
QY 899 CCCTGGGCTTCTGCTGCTCCATTCATGAGAGAGCGGCTGTGTGCTGAGACCTGAA 958  
DB 439 CCTTCTGCTGCTGCTCCATTCATGAGAGAGCGGCTGTGTGCTGAGACCTGAA 381  
QY 959 GTGAGAGACCCCAAGCAGTGAACAGCTTCTTGAG-CCCATGAGAGAGTGAAGAGAG 1017  
DB 380 GTGAGAG-NCAGAGAGCAGTGAAGAGAGCTTCTTGAGAGAGAGTGAAGAGAG 322  
QY 1018 CCTGTGTATGAGAGAGTCCCTTGGGCTTAACTCATGTTCCGCAATGAGTCCGCTGT 1077  
DB 321 CCTGTGTATGAGAGAGAGTCCCTTGGGCTTAACTCATGTTCCGCAATGAGTCCGCTGT 263  
QY 1078 ACCTCATTCATGAGTCCCATTTGTACTGTACAGAGTGTGAGAGAACTGAATGAGAGCA 1137  
DB 262 ACCTCATTCATGAG-CCCAATTTGTACTGTACAGAGTGTGAGAGAACTGAATGAGCA-NC 205  
QY 1138 AGGCTCTTCCGCAACCCCAAGATGAGACAGTCTCTGAGCTGCACTGAGAGGCTT 1197  
DB 204 AGGCTCTTCCGCAACCCCAAGATGAGACAGTCTCTGAGCTGCACTGAGAGGCTT 146  
QY 1198 TCTGATGATGAGTGTGAGTCCCTGAGTCCCTGAGTCCCTGATTAAGACATTTGCTT 1257  
DB 145 TCTGATGATGAGTGTG-CTGAGGCTGAGTCCCTGAGTCCCTGATTAAGACATTTGCTT 89  
QY 1258 CTTCGCTTCTCACTGAGAGGCTTGTGAGATGAGTGTGAGTCCCTGAGTCCCTGAGT 1317  
DB 88 CTTCGCTTCTCACTGAG-GGAGCTTGTGAGATGAGTGTG-CTGAGTCCCTGAGTCCCTGAGT 33  
QY 1318 TCTCATTAATTAAGACAGATGCTTCTTCACTT 1349  
DB 32 TCTCATTAATTAAGACAGATGCTTCTTCACTT 1

RESULT 13  
US-09-960-352-10323  
; Sequence 10323, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Ningbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 10323  
; LENGTH: 420  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 44-LIB34-041-Q1-E1-C8  
US-09-960-352-10323

Query Match 9.4%; Score 298.4; DB 10; Length 420;  
Best Local Similarity 81.9%; Pred. No. 3.6e-88;  
Matches 344; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 663 AGCCGAGATGCTGAGATGAGTCCCATTCATGAGTGTGAGAGAGGCTTCTGAGCA 722  
DB 1 AGCCGAGATGCTGAGATGAGTCCCATTCATGAGTGTGAGAGAGGCTTCTGAGCA 60  
QY 723 GTGAGATGCTGCTTTTCTGAGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGTAT 782  
DB 61 GTGAGATGCTGCTTTTCTGAGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGTAT 120

QY 783 GTCTCTGACCAAGAGAGGCTATACCTAGTAAGGCTTATCCAGAGGCTGAGAG 842  
DB 121 ATTCTCTGACCAAGAGAGGCTATACCTAGTAAGGCTTATCCAGAGGCTGAGAG 180  
QY 843 GAAGTGGAGACCCCTGATGAGATTCCTGAGACTGTGAGAGCGGCTTTATCTGCT 902  
DB 181 GAATTTGGAGAGGCTGATGAGGCTGCTGCTTCAATTCGTGAGAGAGGCTTTATCTGCT 240  
QY 903 GGCTCTTCTGCTGCTGCTATTCATGAGAGAGAGGCTGAGAGGCTGAGAGGCTGAG 962  
DB 241 GAATTTCTGCTGCTGCTGCTATTCATGAGAGAGAGGCTGAGAGGCTGAGAGGCTGAG 300  
QY 963 GGAGCCCAAGCAGTGAACAGAGCTTCTTGAGGCTTCAAGAGAGTGAAGAGTGAAGAG 1022  
DB 301 GGAGCTCAAGCAGCAGTGAACAGAGCTTCTTGAGGCTTCAAGAGAGTGAAGAGTGAAG 360  
QY 1023 TGTATGAGAGAGTCCCTTGGGCTTAACTCATGTTCCGCAATGAGTCCGCTTTGACT 1082  
DB 361 TGTATGAGAGAGTCCCTTGGGCTTAACTCATGTTCCGCAATGAGTCCGCTTTGACT 420

RESULT 14  
US-09-960-352-7040  
; Sequence 7040, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Ningbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 7040  
; LENGTH: 408  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 30-LIB34-026-Q1-E1-H5  
US-09-960-352-7040

Query Match 8.8%; Score 280; DB 10; Length 408;  
Best Local Similarity 80.4%; Pred. No. 4.9e-82;  
Matches 328; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 767 GGAGCACCAGATATATGCTTCTGACCAAGAGAGGCTTATACCTAGTAAGGCTTATCC 826  
DB 1 GGAGTACCAGATATATATCTTCTGACCAAGAGAGGCTTATACCTAGTAAGGCTTATCC 60  
QY 827 GAAGCGCTGAGAGAGAGAGTGGAGACCCCTCATGAGATTCCTGAGTCTGTGATGC 886  
DB 61 AAAGCAGCTGAGAGAGAGATTTGGAGCCCTCATGAGGCTCTTCAATTCGTGTGATGC 120  
QY 887 GGCTTTATCTGCTGCTGCTTCTGCTGCTCATATCATGAGAGAGGCTGAGAGGCTGAG 946  
DB 121 AGCTTTATCTGCTGCTGCTGCTTCTGCTGCTCATATCATGAGAGAGGCTGAGAGGCTGAG 180  
QY 947 GCTGACCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGTCTTCTGAGAGAGAGAG 1006  
DB 181 GCTGACCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGTCTTCTGAGAGAGAGAG 240  
QY 1067 TCCGCTTGTACCTCATGAGTGTGAGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGT 1126  
DB 301 TCTGAGCTTGTACCTCATGAGTGTGAGAGAGAGAGAGAGAGTGTGAGAGAGAGAGT 360  
QY 1127 GAATGAGCAGAGGCTTCCGCAACCCAGATGAGAGAGTCTTCTT 1174





Db	121	CTGGCTCTGGGCATTGGCACCCCTCTTCTCTCCGACTAAGTCCCATGGGAATTTGCTGA	180
Qy	125	AGCGCAGACCAAGCCAGACCCAGACGTGACTGACCGCTCTCAGATGGCTGAGCTTTGA	184
Db	181	AGGGCAGACCAAGCCAGACCCAGACGTGACTGACCGCTCTCAGATGGCTGAGCTTTGA	240
Qy	185	TGCTACCAACCTTGATGACAATGGAACCATGCTGTTTTTAAAGGGAGTTGTGTGAA	244
Db	241	TGCTACCAACCTTGATGACAATGGAACCATGCTGTTTTTAAAGGGAGTTGTGTGAA	300
Qy	245	GAGTCACAAATGGGACCCGGAGTTAATCTCAGAGAGATGGAGAATTTCCCGACCTGT	304
Db	301	GAGTCACAAATGGGACCCGGAGTTAATCTCAGAGAGATGGAGAATTTCCCGACCTGT	360
Qy	305	GGATGCTCATTTCCCTCAAGTCAACAAGTGTCTTTCTGATCAAGGGGACAAAGTCTG	364
Db	361	GGATGCTCATTTCCCTCAAGTCAACAAGTGTCTTTCTGATCAAGGGGACAAAGTCTG	420
Qy	365	GGTATACCTCCTGAAAAAGAGGAAGATPACCAAGTTGCTCCAAATGAATTTC	424
Db	421	GGTATACCTCCTGAAAAAGAGGAAGATPACCAAGTTGCTCCAAATGAATTTC	480
Qy	425	TGGAATC 431	
Db	481	TGGAATC 487	

RESULT 8  
US-09-918-995-8595  
; Sequence 8555, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:

```

1  APPLICANT: Hyseq, Inc.
2  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
3  TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
4  FILE REFERENCE: 20411-756
5  CURRENT APPLICATION NUMBER: US/09/918,995
6  CURRENT FILING DATE: 2001-07-30
7  PRIOR APPLICATION NUMBER: US/09/235,076
8  PRIOR FILING DATE: 1999-01-20
9  NUMBER OF SEQ. ID NOS: 38054
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 8595
12 LENGTH: 422
13 TYPE: DNA
14 ORGANISM: Homo sapiens
15 US-09-918-995-8595

```

Query Match	Score	DB	Length
13.2%	421	11	422

```

Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	2	TCTGAGCTCAGCAGGCGTAAAGGTACTGGAGACACCGTTGACATGGGGTTTGGAGCCT	61
Db	1	TCTGACGCTCAGCAGTGGCTTAAGGTACTGGAGACACCGTTGACATGGGGTTTGGAGCCT	60
QY	62	ATGCTGTCTCTGAGCAATTTGCCACCCCTCTTCTCCGACTAAGTCCCATGGGAATGTTGC	121
Db	61	ATGCTGTCTCTGAGCAATTTGCCACCCCTCTTCTCCGACTAAGTCCCATGGGAATGTTGC	120
QY	122	TGAAGGCGAGACCAAGCCAGACCCAGACGATGACTGAAAGCTCTCAGATGGCGTGAAGCTT	181
Db	121	TGAAGGCGAGACCAAGCCAGACCCAGACGATGACTGAAAGCTCTCAGATGGCGTGAAGCTT	180
QY	182	TGATGCTTACCAACCCCTGATGACAAATGGAACCAATGCTGTTTTTAAAGGGGAGTTTGTGTG	241
Db	181	TGATGCTTACCAACCCCTGATGACAAATGGAACCAATGCTGTTTTTAAAGGGGAGTTTGTGTG	240
QY	242	GAAAGTCACAAAATGGGACCGGGATTTAATCTCAGAGAGATGGAGAAATTTTCCCGACGCC	301
Db	241	GAAAGTCACAAAATGGGACCGGGATTTAATCTCAGAGAGATGGAGAAATTTTCCCGACGCC	300

Qy	302	TTGTGATNCTGATTCGGTCAAGTCACAAAGTGTCTTTGTGATCAAGGGGACAAAGT	361
Db	301	TTGTGATNCTGATTCGGTCAAGTCACAAAGTGTCTTTGTGATCAAGGGGACAAAGT	360
Qy	362	CTGGGATATACCTCTCTGAAGAAGAGAAAGATACCCAAGTGTCTCAAGATGAATT	421
Db	361	CTGGGATATACCTCTCTGAAGAAGAGAAAGATACCCAAGTGTCTCAAGATGAATT	420
Qy	422 T 422		
Db	421 T 421		

RESULT 9  
US-09-918-995-31395

```

; Sequence 31395, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

```

```

? TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
? TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
? FILE REFERENCE: 20411-756
? CURRENT APPLICATION NUMBER: US/09/918,995
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: US/09/235,076
? PRIOR FILING DATE: 1999-01-20
? NUMBER OF SEQ ID NOS: 38054
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 31395
? LENGTH: 492
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(492)
? OTHER INFORMATION: n = A,T,C or G
? US-09-918-995-31395

```

Query Match	Score	DB	Length
13.1%	417	11	492

Matches 434; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Qy	5	GCACCTCAGCATGGCTTAGGGTACTGGGAGACCCGTTGCACCTGGGGTTGTGGACCTATG	64
Dd	46	GCACCTCAGCATGGCTTAGGGTACTGGGAGACCCGATGCACCTGGGGTTGTGGACCTATG	105
Qy	65	CTGTCTCTGGCCATTTGCCACCCCTCTTCTCCGACTAGTGCCTCAATGGAAATGTTGCTGA	124
Dd	106	CTGTCTCTGGCCATTTGCCACCCCTCTTCTCCGACTAGTGCCTCAATGGAAATGTTGCTGA	165
Qy	125	AGGGGAGAACCAAGCCACACCCGAGCGTGAACGCTGCCTCAATGGCTGGACCTTTGA	184
Dd	166	AGGGGAGAACCAAGCCACACCCGAGCGTGAACGCTGCCTCAATGGCTGGACCTTTGA	225
Qy	185	TGCTACCAACCTGGATGACAAATGGAACATCTGTTTTTAAAGGGAGTTGTGTGAA	244
Dd	226	TGCTACCAACCTGGATGACAAATGGAACATCTGTTTTTAAAGGGAGTTGTGTGAA	285
Qy	245	GAGTCACAAATGGGACCGGGAGTTAATCTCAGAGAGATGGAGAAATTTCCCAAGCCTGT	304
Dd	286	GAGTCACAAATGGGACCGGGAGTTAATCTCAGAGAGATGGAGAAATTTCCCAAGCCTGT	345
Qy	305	GGATGCTGCATTCGGTCAGAGTCAACAAGTGTCTTTCTGATCAAGGGGGCAAAAGTCTG	364
Dd	346	GGATGCTGCATTCGGTCAGAGTCAACAAGTGTCTTTCTGATCAAGGGGGCAACAAGTCTG	405
Qy	365	GGTATTAACCTTCCTGAAAAAAGAGGAAGATACCAAAAGTTGCTCCAAAGTGAATTTCC	424
Dd	406	AGTATTAACCTTCCTGAAAAAAGAGGAGACAGGATACCCAAAAGTGTCTCCAAAGTGAATTTAC	465
Qy	425	TGGAATCCCATCCCACTGAGTGAAGC	451
Dd	466	TGG-ATCCCAATCCCACTGAGTGAAGC	491

SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 84  
 LENGTH: 1516  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 US-10-175-523-84

Query Match 18.1%; Score 576.8; DB 15; Length 1516;  
 Best Local Similarity 67.3%; Pred. No. 1.1e-180;  
 Matches 999; Conservative 0; Mismatches 257; Indels 228; Gaps 4;

```

QY 1 CTCTGACGCTCAGCATGCTAGAGGTAAGTCTGGAGGACCCGTTGACCTGGAGGTTTGAGGCC 60
DB 41 CTTTGACGCTCGCATGCTAGAGGTAAGTCTGGAGGACCCGTTGACCTGGAGGTTTGAGGCC 100
QY 61 TATGCTGCTCTGGCCATGCTAGGACCCCTCTTCTCCGACTAGTCCCATGGGAATGTTG 120
DB 101 TGTGCTGCTCTGGCCATGCTAGGACCCCTCTTCTCCGACTAGTCCCATGGGAATGTTG 160
QY 121 CTGAGGCGAGACCCAGACCCAGACCCGTAAGTGAAGCTGCTGATGGCTGAGACT 180
DB 161 GTGAAATGGAGCCAGACCCAGACTCAGATGTAATGCAACCTGCTCAGATGCTGAGCT 220
QY 181 TTGATGCTACACCCCTGATGACATGGAACCATGCTGTTTAAAGGGAATGTTGCT 240
DB 221 TTGACGCTACACCCATGATGACATGGAACCATGCTGTTTAAAGGGAATGTTGCT 280
QY 241 GGAAGATCAACAATGGAGCCGGAGTTAATCTCAGAGATGGAAGATTTCCCGAGCC 300
DB 281 GGAAGGATCACTAGGAGATCGGAGTTAATCTCAGAGATGGAAGATTTCCCGAGCT 340
QY 301 CTGAGATGCTGATTCGCTCAAGTCAACAGTGTCTTTCTGATCAAGGGGAGCAAG 360
DB 341 CAGTGGATGCTGATTCGCTCAAGTGTCTTTCTGATCAAGGGGAGCAAG 397
QY 361 TCTGGGATATACCTCTGAAAGAGAGAAAGGATACCAAGATTGCTCCAGATGAT 420
DB 398 TCTGGGATATACCTCTGAAAGAGAGAAAGGATACCAAGATTGCTCCAGATGAT 457
QY 421 TTCTGGAATTCCTATCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 458 CTCTGGAATTCCTATCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 517
QY 481 CTGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 508
DB 518 GTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 577
QY 509 ----- 508
DB 578 CCAAAAGAAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637
QY 509 ----- 508
DB 638 AACCTACTACTGCTTCCAGGTAACAAGTTCTGAGATTTAACCCGTCACAGAGAG 697
QY 509 ----- 508
DB 698 TGCTTCCCAATACCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 757
QY 509 -----TGACACAGAGATGAGAGCTGGCCATGAGAAAGTATCCCATGAGCTGATGATA 564
DB 758 ATGCTAAATCTAAAGAAATGGAATGCTCTCAATGGAATGAGACCA-----TCTATGAT 811
QY 565 TGCGCTGATGAGCCATCTAGTCTTCTCTGATGATGATGATGATGATGATGATGAT 624
DB 812 CGCTGTAAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 871
QY 625 ATGCTTCAAGTGAAGCACTACTGAGCTGAGACACAGCGGAGATGAGCTGAGATGCT 684
DB 872 ATGCTTCAAGTGAAGCACTACTGAGCTGAGACACAGCGGAGATGAGCTGAGATGCT 931
QY 685 GAGCCATGCTCAATCAATGAGGCGCCAGAGCTCTTGAAGAGATGCTGCTTTTCTGAG 744

```

```

DB 932 GGCCATTCCTCATCATGCTGCCCCAGGGTCTTTCAGAGTGAATGCTGCTTTTCTGAG 991
QY 745 AAGAAATCTATCTATGCTGCTGAGGAGCCAGGATATATGCTTCTGCAAGAAAGGAGCT 804
DB 992 ATGAGAAATCTATCTATGCTGCTGAGGAGCCAGGATATATGCTTCTGCAAGAAAGGAGCT 1051
QY 805 ATACCTTAAAGCGGTTATCCGAAGCGGCTGAGAGAGAGATGCGGACCCCTCATGAGGA 864
DB 1052 ATACCTTAAAGCGGTTATCCGAAGCGGCTGAGAGAGAGATGCGGACCCCTCATGAGGA 1111
QY 865 TTAATCTGATCTTGTGATGAGGCGCTTATATGCTGCTGCTGCTGCTGCTGCTGCT 924
DB 1112 TCAGCTTATATCAATAGATGAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1171
QY 925 TGACAGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984
DB 1172 CATGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1231
QY 985 AGCTTCTTGGCCCATGAGAAAGTGAAGAGAGCTTGTATGAAAGATCCCTTGGCC 1044
DB 1232 AGCTTCTTGGCCCATGAGAAAGTGAAGAGAGCTTGTATGAAAGATCCCTTGGCT 1291
QY 1045 CTACTCATGCTCTTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104
DB 1292 CTACTCATGCTCTTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1351
QY 1105 GCTACAGTATGAGGAAATGGAATGAGCCAGGCGCTTCCGCAACCCAGAAATGTA 1164
DB 1352 GCTACAGTATGAGGAAATGGAATGAGCCAGGCGCTTCCGCAACCCAGAAATGTA 1411
QY 1165 CCAATCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221
DB 1412 ACAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1471
QY 1222 CACTCTGATGCTCTGATATATTAAGACAGATGCTTCTGCTT 1265
DB 1472 CACTCTGATGCTCTGATATATTAAGACAGATGCTTCTGCTT 1515

```

## RESULT 7

US-09-918-995-30828  
 ; Sequence 30828, Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-756  
 ; CURRENT FILING DATE: 2001-07-30  
 ; PRIOR FILING DATE: 1999-01-20  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 30828  
 ; LENGTH: 488  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(488)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-918-995-30828

Query Match 13.4%; Score 427; DB 11; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-131;  
 Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 5 GCAGCTCAGCATGCTAGGTAAGTCTGGAGGACCCGTTGACCTGGAGGTTTGAGGCC 64
DB 61 GCAGCTCAGCATGCTAGGTAAGTCTGGAGGACCCGTTGACCTGGAGGTTTGAGGCC 120
QY 65 CTGCTCTGCGCATGCTGCAACCCCTCTCTCTGCACTAGTCCCATGGAATGTTGCTGA 124

```

NUMBER OF SEQ ID NOS: 308  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 39  
 LENGTH: 1516  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (55)..(1437)  
 OTHER INFORMATION:  
 US-10-316-253-39

Query Match 18.1%; Score 576.8; DB 13; Length 1516;  
 Best Local Similarity 67.3%; Pred. No. 1,1e-180;  
 Matches 999; Conservative 0; Mismatches 257; Indels 228; Gaps 4;

QY 1 CTCTGACGTACAGATGCTAGAGGTACTGGAGACCCGTTGACCTGGGGTTGTGAGCC 60  
 DB 41 CTCTGACGTACAGATGCTAGAGGTACTGGAGACCCGTTGACCTGGGGTTGTGAGCC 100  
 QY 61 TATGCTGCTCTGGCCATGTCACCCCTCTTCTCCGACTAGTGGCCATGGGAATGTTG 120  
 DB 101 TATGCTGCTCTGGCCATGTCACCCCTCTTCTCCGACTAGTGGCCATGGGAATGTTG 160  
 QY 121 CTGAAGGCGAGACCAAGCCAGACCCGACGTAAGTGAAGCTGCTGATGAGTGGAGCT 180  
 DB 161 GTGAAATGGGACCAAGCCAGACCCGACGTAAGTGAAGCTGCTGATGAGTGGAGCT 220  
 QY 181 TTGATCTACACCTCTGATGACATGGAACCATGCTGTTTTTAAAGGGAGTTGTGT 240  
 DB 221 TTGATCTACACCTCTGATGACATGGAACCATGCTGTTTTTAAAGGGAGTTGTGT 280  
 QY 241 GGAAGGTCACCAATGGGACCGGAGTATCTCAAGAGATGGAATTTCCCAAGCC 300  
 DB 281 GGAAGGTCACCAATGGGACCGGAGTATCTCAAGAGTGAATTTCCCAAGCT 340  
 QY 301 CTGTGATGCTGATTCCTGCAAGTCAACAAGTCTTCTGATCAAGGGGACCAAG 360  
 DB 341 CAGTGATGCTGATTCCTGCAAGTCAACAAGTCTTCTGATCAAGGGGACCAAG 397  
 QY 361 TCTGGGTATACCTCTCTGAAAGAGAGAAAGATACCAAAAGTTCTCAAGATGAT 420  
 DB 398 TCTGGGTATACCTCTCTGAAAGAGAGAAAGATACCAAAAGTTCTCAAGATGAT 457  
 QY 421 TTCTGTGAATCCCATCCCACTGATGACACTGTGGAATGCAACCGTGGAGAAATGTCAG 480  
 DB 458 CTCTGTGAATCCCATCCCACTGATGACACTGTGGAATGCAACCGTGGAGAAATGTCAG 517  
 QY 481 CTGAAGGCGTCTCTTCTTCAAGGCCA----- 508  
 DB 518 GTGAAGGTCCTCTTCTTCAAGGTAACCGCAAGTGTCTGGGACTTTCACAAAGAA 577  
 QY 509 ----- 508  
 DB 578 CCAAAAGAAAGTTCCTGCTGCTGTGGAATGCACTGGGCTTGAAGTGGCTTG 637  
 QY 509 ----- 508  
 DB 638 AACGTACTACTGCTTCCAGGGTAACAAGTTCTGAGATTTAAACCCGTCACAGAGAGG 697  
 QY 509 ----- 508  
 DB 698 TGCTCCCAAGTACCTCTGGAATGCCGTAATCACTTATCTGCTGCTGGAGAGGCC 757  
 QY 509 -----TGGAACAAGAAATGGGATGGCCATGGGAAAGTAACCAATGAGCCCTGAGATA 564  
 DB 758 ATGTAACAACTAAAGAAATGGAATGCTCATGGAATAGCAACCA-----TCTATGATTT 811  
 QY 565 TGGCTGTAGCCCACTATGATTTGTTCTGCACTGACGTCTGAACCAATGTTGCCACT 624  
 DB 812 CGGCTGTAGCAAGATCTCGGCTGTGCACTGCTGTCTGACATCGAGGTCACCT 871  
 QY 625 ATGCTTCAAGTGGAGCCCACTACTGGGCTGTGACACCAAGCGGGAATGGTGGCATAGCT 684

DB 872 ATGCTTCAAGTGGCTCCCACTACTGGGCTGTGACTGCAAGCGTGAATGGGTGATAGCT 931  
 QY 685 GGGCCATTGCTATCATGATGGCCCAAGGTCCTTACAGATGATGCTGCTTTCTGAGG 744  
 DB 932 GGGCCATTGCTATCATGATGGCCCAAGGTCCTTACAGATGATGCTGCTTTCTGAGG 991  
 QY 745 AAGAAATCTATCATGATGGCCCAAGGTCCTTACAGATGATGCTGCTTTCTGAGAGGCT 804  
 DB 992 ATGAAATCTATCATGATGGCCCAAGGTCCTTACAGATGATGCTGCTTTCTGAGAGG 1051  
 QY 805 ATACCTTATGACCGGTTATCCGAACCGCTGGAAGAAAGTGGAGACCCCTCATGGA 864  
 DB 1052 ATACCTTATGATGTTATTCGAAGCGCTGGAAGAAAGTGGAGACCCCTCGGGA 1111  
 QY 865 TTATCTGATCTGTGATGATGGGCTTTATGCTGCTGCTGCTTCTGATGATCA 924  
 DB 1112 TCAGCTTATATCAATGATGACGCTTTCTGCTGCTGCTTCTTCAAGCTTACGTC 1171  
 QY 925 TGGCAGAGCGGCGCTGTGGCTGAGCCTGGAAGCAGAGCCCAAGCCAGTGGACAG 984  
 DB 1172 CATCAGAGCGGCGCTGTGGCTGAGCCTGGAAGCAGAGCCCAAGCCAGTGGACAG 1231  
 QY 985 AGCTTCTTGGCCCATGAGAGGTAGAGCGAGCCTTGTGTATGAAAGTCCCTTGGCC 1044  
 DB 1232 AGCTTCTTGGCCCATGAGAGGTAGAGCGAGCCTTGTGTATGAAAGTCCCTTGGCT 1291  
 QY 1045 CTATCTATGTTCCGCAATGATCCGCTGTTACTCATTCATGATGCCAATTTGACT 1104  
 DB 1292 CCTACTCATGCTTCTTCAATGATGATGCTTCTTCTTATCAATGAGGCCAATTTTACT 1351  
 QY 1105 GCTACAGTATGAGGAAATGAAATGCAAGCCAGCCCTCCGCAACCCAGAAATGTA 1164  
 DB 1352 GCTATGCAATATGAAATGAAATGCAAGCCAGCCCTCCGCAACCCAGAAATGTA 1411  
 QY 1165 CCAGTCTCTGGGCTGCACTACTGAGGGCC---TTCTGACATGATGCTGGCTGGCC 1221  
 DB 1412 AACGATCTTGGCTGCACTACTGAAATGAAAGCCGTATGGAATTTAGCCAGCCCA 1471  
 QY 1222 CACTCTATGTTCTCATTAATAAGACAGATTTCTTCTGCTT 1265  
 DB 1472 CACTCTATGTTCTCATTAATAAGACAGATTTCTTCTGCTT 1515

RESULT 6  
 US-10-175-523-84  
 ; Sequence 84, Application US/10175523  
 ; Publication No. US20030096264A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brockman, Jeffrey  
 ; APPLICANT: Evans, David  
 ; APPLICANT: Hook, Derek  
 ; APPLICANT: Klimczak, Leszek  
 ; APPLICANT: Laeng, Pascal  
 ; APPLICANT: Paley, Michael  
 ; APPLICANT: Rajan, Prithi  
 ; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)  
 ; FILE REFERENCE: 3235/10795-US3  
 ; CURRENT APPLICATION NUMBER: US/10/175,523  
 ; PRIOR FILING DATE: 2002-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/299,151  
 ; PRIOR FILING DATE: 2001-06-18  
 ; PRIOR APPLICATION NUMBER: US 60/317,828  
 ; PRIOR FILING DATE: 2001-09-07  
 ; PRIOR APPLICATION NUMBER: US 60/325,150  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/333,047  
 ; PRIOR FILING DATE: 2001-11-14  
 ; PRIOR APPLICATION NUMBER: US 60/349,936  
 ; PRIOR FILING DATE: 2002-01-18  
 ; PRIOR APPLICATION NUMBER: US 60/361,834  
 ; PRIOR FILING DATE: 2002-03-04  
 ; NUMBER OF SEQ ID NOS: 197

/ CURRENT APPLICATION NUMBER: US/10/105, 891  
 / CURRENT FILING DATE: 2002-03-25  
 / PRIOR APPLICATION NUMBER: 09/668,317  
 / PRIOR FILING DATE: 2000-09-22  
 / PRIOR APPLICATION NUMBER: 09/552,929  
 / PRIOR FILING DATE: 2000-04-18  
 / NUMBER OF SEQ ID NOS: 91  
 / SOFTWARE: pc\_Fl\_genes Version 2.0  
 / SEQ ID NO 19  
 / LENGTH: 1631  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: (71) .. (1459)  
 / US-10-105-891-19

Query Match 35.5%; Score 1129.8; DB 15; Length 1631;  
 Best Local Similarity 86.3%; Pred. No. 0;  
 Matches 1354; Conservative 0; Mismatches 2; Indels 213; Gaps 1;

QY 1 CTCTGAGCTCAGCATGCTAGGTAAGTGGAGACCCCTTGCATCTGGGTTGTGAGCC 60  
 DB CTCTGAGCTCAGCATGCTAGGTAAGTGGAGACCCCTTGCATCTGGGTTGTGAGCC 116  
 QY 61 TATGCTGCTCTGAGCATGCTAGGTAAGTGGAGACCCCTTGCATCTGGGTTGTGAGCC 120  
 DB 117 TATGCTGCTCTGAGCATGCTAGGTAAGTGGAGACCCCTTGCATCTGGGTTGTGAGCC 176  
 QY 121 CTGAGGCGGAGACCAAGCCAGACCGTGAAGTGAAGCTGCTGATGCTGAGCT 180  
 DB 177 CTGAGGCGGAGACCAAGCCAGACCGTGAAGTGAAGCTGCTGATGCTGAGCT 236  
 QY 181 TTGATGCTACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 237 TTGATGCTACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 296  
 QY 241 GGAAGAGTCAAGATGGAGACCGGAGATTATCTCAGAGAGTGAAGATTTCCAGCC 300  
 DB 297 GGAAGAGTCAAGATGGAGACCGGAGATTATCTCAGAGAGTGAAGATTTCCAGCC 356  
 QY 301 CTGTGATGCTGATTCCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360  
 DB 357 CTGTGATGCTGATTCCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 416  
 QY 361 TCTGGTATATCCCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 DB 417 TCTGGTATATCCCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 476  
 QY 421 TTCTGGAATCCCATCCCACTGGAATGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480  
 DB 477 TTCTGGAATCCCATCCCACTGGAATGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 536  
 QY 481 CTGAGGCGCTCTCTTCTCA----- 502  
 DB 537 CTGAGGCGCTCTCTTCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 596  
 QY 503 ----- 502  
 DB 597 CCATGAAGAGGCTTCTGAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 656  
 QY 503 ----- 502  
 DB 657 GCCGCTACTAGCTTCAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 716  
 QY 503 -----AGGCC 507  
 DB 717 TGCCTCCAGAGTACCCCGGAGATGCGAGACTACTTATGATCCCTGCGCTGAGAGAGCC 776  
 QY 508 ATGAGACAGAGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 567  
 DB 777 ATGAGACAGAGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 836

QY 568 GGTGAGCCCACTAGTCTTGTCTGCACTGACGTCTGCAACAGTGTGCCACTATG 627  
 DB 837 GGTGAGCCCACTAGTCTTGTCTGCACTGACGTCTGCAACAGTGTGCCACTATG 896  
 QY 628 CTTTCAAGTGAAGCCCACTAGTCTTGTCTGCACTGACGTCTGCAACAGTGTGCCACTATG 687  
 DB 897 CTTTCAAGTGAAGCCCACTAGTCTTGTCTGCACTGACGTCTGCAACAGTGTGCCACTATG 956  
 QY 688 CCAATGCTCATGAGTGGGCGGCGGCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 747  
 DB 957 CCAATGCTCATGAGTGGGCGGCGGCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1016  
 QY 748 AAAAATCTATGCTGACAGGAGCCAGATATGCTTCTGCAACAGTGTGCCACTATG 807  
 DB 1017 AAAAATCTATGCTGACAGGAGCCAGATATGCTTCTGCAACAGTGTGCCACTATG 1076  
 QY 808 CCTTGAAGCGGTTATCCGAGCGGCTGAGAGAGAGTGGGAGCCCTCATGAGATTA 867  
 DB 1077 CCTTGAAGCGGTTATCCGAGCGGCTGAGAGAGAGTGGGAGCCCTCATGAGATTA 1136  
 QY 868 TCTTGAAGCTGAGTGGGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 927  
 DB 1137 TCTTGAAGCTGAGTGGGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1196  
 QY 928 CAGAGCGGCGGCTGAGTGGGCTGAGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 987  
 DB 1197 CAGAGCGGCGGCTGAGTGGGCTGAGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1256  
 QY 988 TTCTTGGCGGCGGCTGAGTGGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1047  
 DB 1257 TTCTTGGCGGCGGCTGAGTGGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1316  
 QY 1048 ACTCATGTTCCGCGAATGCTGAGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1107  
 DB 1317 ACTCATGTTCCGCGAATGCTGAGCGGCTTATGCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGGCTTCTGAGCGG 1376  
 QY 1108 ACAGTATGAGAGAACTGAATGAG 1167  
 DB 1377 ACAGTATGAGAGAACTGAATGAG 1436  
 QY 1168 GTCTCTGAGGCTGACTCACTGAG 1227  
 DB 1437 GTCTCTGAGGCTGACTCACTGAG 1496  
 QY 1228 CTAAGTCTCATTAATAAAG 1287  
 DB 1497 CTAAGTCTCATTAATAAAG 1556  
 QY 1288 TGAAGTGGGCTGAGCGGCGGCTTCAATTAATAAAG 1347  
 DB 1557 TGAAGTGGGCTGAGCGGCGGCTTCAATTAATAAAG 1616  
 QY 1348 TTGAATCAA 1356  
 DB 1617 TTGAATCAA 1625

RESULT 5  
 US-10-316-253-39  
 / Sequence 39, Application US/10316253  
 / Publication No. US20030162706A1  
 / GENERAL INFORMATION:  
 / APPLICANT: The Procter & Gamble Company  
 / APPLICANT: Peters, Kevin  
 / APPLICANT: Thompson, Larry  
 / APPLICANT: Wang, Feng  
 / APPLICANT: Greis, Kenneth  
 / TITLE OF INVENTION: Angiogenesis Modulating Proteins  
 / FILE REFERENCE: 885M  
 / CURRENT APPLICATION NUMBER: US/10/316,253  
 / CURRENT FILING DATE: 2002-12-10  
 / PRIOR APPLICATION NUMBER: US 60/355,295  
 / PRIOR FILING DATE: 2002-02-08

PRIOR FILING DATE: 2000-09-22  
 PRIOR APPLICATION NUMBER: 09/552,929  
 PRIOR FILING DATE: 2000-04-18  
 NUMBER OF SEQ ID NOS: 91  
 SOFTWARE: pc\_fl\_genes Version 2.0  
 SEQ ID NO 19  
 LENGTH: 1631  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (71)..(1459)  
 US-10-125-237-19

Query Match 35.5%; Score 1129.8; DB 15; Length 1631;  
 Best Local Similarity 86.3%; Pred. No. 0;  
 Matches 1354; Conservative 0; Mismatches 2; Indels 213; Gaps 1;

QY 1 CTCCTGACGCTCAGCATGGCTAGGGTACTGGAGCACCCTGTCAGTGGGGTTGTGAGCC 60  
 DB 57 CTCCTGACGCTCAGCATGGCTAGGGTACTGGAGCACCCTGTCAGTGGGGTTGTGAGCC 116  
 QY 61 TATGCTGCTCTGGCCATTGCCACCCCTCTTCTCCGACTAGTGCCTATGGGAATGTTG 120  
 DB 117 TATGCTGCTCTGGCCATTGCCACCCCTCTTCTCCGACTAGTGCCTATGGGAATGTTG 176  
 QY 121 CTGAAGCGGAGCAGCAGCCAGCCAGCAGTGAAGCTGACCTGTCAGATGGCTGAGCT 180  
 DB 177 CTGAAGCGGAGCAGCAGCCAGCCAGCAGTGAAGCTGACCTGTCAGATGGCTGAGCT 236  
 QY 181 TTGATGCTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 237 TTGATGCTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 296  
 QY 241 GGAAGAGTCAAAATGGAGCCGGAGTTAATCTCAGAGATGAGAAATTTCCCAAGCC 300  
 DB 297 GGAAGAGTCAAAATGGAGCCGGAGTTAATCTCAGAGATGAGAAATTTCCCAAGCC 356  
 QY 301 CTGTGATGCTGATTCCTGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 DB 357 CTGTGATGCTGATTCCTGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 416  
 QY 361 TCTGGTATACCTCCTGAAAGAGAGAGATACCCAAAGTTGCTCAAGATGAT 420  
 DB 417 TCTGGTATACCTCCTGAAAGAGAGATACCCAAAGTTGCTCAAGATGAT 476  
 QY 421 TTCTGGAATCCATCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 477 TTCTGGAATCCATCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 536  
 QY 481 CTGAAGCGCTCTCTTCTTCCA----- 502  
 DB 537 CTGAAGCGCTCTCTTCTTCCA----- 596  
 QY 503 ----- 502  
 DB 597 CCATGAAGAGCGTTCTGCGCAGCTGTGGGAATGCTCTCTGCTTGAATGCTGG 656  
 QY 503 ----- 502  
 DB 657 GCCGCTACTACTGCTTCCAGGGTAAACCAATTCTGCGCTTGAACCTGTGAGGAGAG 716  
 QY 503 -----AGGCC 507  
 DB 717 TGCCTCCAGGTACCCGCGGGAATGTCGAGACTACTTATGCTGCTGCTGCGAGAGGCC 776  
 QY 508 ATGAACACAGGAATGGGACTGGCCATGGGAAAGTATCCCAATGGCCCTGAGATATATGC 567  
 DB 777 ATGAACACAGGAATGGGACTGGCCATGGGAAAGTATCCCAATGGCCCTGAGATATATGC 836  
 QY 568 GCTGTAGCCCACTATAGTCTTGTCTGCACTGACGTGTGAACAACATGATGCTGACCTATG 627  
 DB 837 GCTGTAGCCCACTATAGTCTTGTCTGCACTGACGTGTGAACAACATGATGCTGACCTATG 896

QY 628 CCTTCAGTGGGAGCCCTAATGCGCTGTGAGACACAGCCGGAGATGGCTGATAGTGGC 687  
 DB 897 CCTTCAGTGGGAGCCCTAATGCGCTGTGAGACACAGCCGGAGATGGCTGATAGTGGC 956  
 QY 688 CCATTCCTCATCATGAGGAGCCAGGGTCTCTTCAAGCAGTGAATGCTGCTTCTTCTGGAG 747  
 DB 957 CCATTCCTCATCATGAGGAGCCAGGGTCTCTTCAAGCAGTGAATGCTGCTTCTTCTGGAG 1016  
 QY 748 AAAAATCTATCTGTGTCAGAGGACCCAGGATATATGCTTCTGACAAAGAGAGCTATA 807  
 DB 1017 AAAAATCTATCTGTGTCAGAGGACCCAGGATATATGCTTCTGACAAAGAGAGCTATA 1076  
 QY 808 CCTTAGTAAGCGGTTATCCGAAAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGATTA 867  
 DB 1077 CCTTAGTAAGCGGTTATCCGAAAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGATTA 1136  
 QY 868 TCCCTGACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 927  
 DB 1137 TCCCTGACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196  
 QY 928 CAGAGCGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 987  
 DB 1197 CAGAGCGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1256  
 QY 988 TTCTTGGCCCTCAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTA 1047  
 DB 1257 TTCTTGGCCCTCAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTA 1316  
 QY 1048 ACTCATGTTCCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107  
 DB 1317 ACTCATGTTCCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1376  
 QY 1108 ACAATGATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167  
 DB 1377 ACAATGATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436  
 QY 1168 GTCTCTGGGCTGCACTGATGAGGAGGCTTGTGATGATGATGATGATGATGATGATGATGAT 1227  
 DB 1437 GTCTCTGGGCTGCACTGATGAGGAGGCTTGTGATGATGATGATGATGATGATGATGATGAT 1496  
 QY 1228 CTAGTTCCTATATATTAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287  
 DB 1497 CTAGTTCCTATATATTAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1556  
 QY 1288 TGAATGCTGCGCTGCGCCCACTTCCCAAGTTTCTCATATTAAGCAGATGATGATGATGAT 1347  
 DB 1557 TGAATGCTGCGCTGCGCCCACTTCCCAAGTTTCTCATATTAAGCAGATGATGATGATGAT 1616  
 QY 1348 TTGAATCAA 1356  
 DB 1617 TTGAATAAAA 1625

RESULT 4  
 US-10-105-891-19  
 ; Sequence 19, Application US/10105891  
 ; Publication No. US20030073099A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Zhou, Ping  
 ; APPLICANT: Xu, Feiyan  
 ; APPLICANT: Zhou, Qiang A.  
 ; APPLICANT: Xu, Aifeng J.  
 ; APPLICANT: Zhang, Jie  
 ; APPLICANT: Wehrman, Tom  
 ; APPLICANT: Wang, Jian-Rui  
 ; APPLICANT: Dimaenac, Radoje T.  
 ; TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and  
 ; FILE REFERENCE: 791CIP2A



Db 1188 TGGTTTCACTGATTAACAAGGACCTTGGTCTGTAACAATCTTCTTTCTTTGATTTGA 11947  
Qy 1397 AAAGTTAGCACTTCTCTCTTTGAGGGTGTGAGCTTCAACAAGGCTGTGAGAAACAAGGA 1456  
Db 11948 AAAGTTAGCACTTCTCTCTTTGAGGGTGTGAGCTTCAACAAGGCTGTGAGAAACAAGGA 12007  
Qy 1457 GGGGACACATAAGGGGCAAACTATCTGTGCGAGATGATTTCTTAGTCCAGATCAATAA 1516  
Db 12008 GGGGACACATAAGGGGCAAACTATCTGTGCGAGATGATTTCTTAGTCCAGATCAATAA 12067  
Qy 1517 CTAGCTCTTTGACAGATCTATCAACATAGTGGGGGAAAGAGAACCAAGATCCGAAAGAG 1576  
Db 12068 CTAGCTCTTTGACAGATCTATCAACATAGTGGGGGAAAGAGAACCAAGATCCGAAAGAG 12127  
Qy 1577 AACAGCTGAGTTTATACAGCAAGTAAAGGTGAGCTAGGACTCTGATTTCACTTGTCTGG 1636  
Db 12128 AACAGCTGAGTTTATACAGCAAGTAAAGGTGAGCTAGGACTCTGATTTCACTTGTCTGG 12187  
Qy 1637 TAGATGGCCACAAACCGGCGCAAGGATGAGAAACAAGGGCTGGGGCACTATGCA 1696  
Db 12188 TAGATGGCCACAAACCGGCGCAAGGATGAGAAACAAGGGCTGGGGCACTATGCA 12247  
Qy 1697 TGTGCAAGAGAGATGTGCTCAGAGTTGTGGGGTGAAGAGTCAATCTGGGGGACTCAAA 1756  
Db 12248 TGTGCAAGAGAGATGTGCTCAGAGTTGTGGGGTGAAGAGTCAATCTGGGGGACTCAAA 12307  
Qy 1757 TTAATGTTCTGGGTGATTCAGATTAACAACATCATGCTTGTGTGTCAGATTAAGCA 1816  
Db 12308 TTAATGTTCTGGGTGATTCAGATTAACAACATCATGCTTGTGTGTCAGATTAAGCA 12367  
Qy 1817 TGACAAGTGAATGAAGTTGAAGTGGGGAAACAGAAATACACAGCTGTGTGTCAGAGG 1876  
Db 12368 TGACAAGTGAATGAAGTTGAAGTGGGGAAACAGAAATACACAGCTGTGTGTCAGAGG 12427  
Qy 1877 AAGCTGAGAGAGAGAGAAAGATGAATGACCAATGAGACCAATTTGGAGAAACAGT 1936  
Db 12428 AAGCTGAGAGAGAGAGAAAGATGAATGACCAATGAGACCAATTTGGAGAAACAGT 12487  
Qy 1937 CCTGGGAGATCTTGTCTGAGAGCTTCAGAGAGCTTGTCTGAGACAGAGATCTGGCTTACA 1996  
Db 12488 CCTGGGAGATCTTGTCTGAGAGCTTCAGAGAGCTTGTCTGAGACAGAGATCTGGCTTACA 12547  
Qy 1997 ATTAGCTCTGGGTATCTGACCAATCTAGACCAAGAAATGTACTGGGCAAGAGAGTTGC 2056  
Db 12548 ATTAGCTCTGGGTATCTGACCAATCTAGACCAAGAAATGTACTGGGCAAGAGAGTTGC 12607  
Qy 2057 AGTGTGGTATTTCTGGTCAATACTGTGAAGGCTTTGGGTTCCAATTTGCTGAC 2116  
Db 12608 AGTGTGGTATTTCTGGTCAATACTGTGAAGGCTTTGGGTTCCAATTTGCTGAC 12667  
Qy 2117 AGCTGTTTAACTGGGAATTTGGGCTTAGACTATAGTAGTATGTCTCAAGCAAGGCCCTTA 2176  
Db 12668 AGCTGTTTAACTGGGAATTTGGGCTTAGACTATAGTAGTATGTCTCAAGCAAGGCCCTTA 12727  
Qy 2177 TTCTCTCAGCTGCTTAAACCAAGCTGAGGCTTGGAGGCTTGGCTTTCAGGCTCAAAA 2236  
Db 12728 TTCTCTCAGCTGCTTAAACCAAGCTGAGGCTTGGAGGCTTGGCTTTCAGGCTCAAAA 12787  
Qy 2237 AATAGCTGAGTTTCCAGAGAGGGGCTTATTTCTAGGCTTCCGTGCTCAGGCTCATTT 2296  
Db 12788 AATAGCTGAGTTTCCAGAGAGGGGCTTATTTCTAGGCTTCCGTGCTCAGGCTCATTT 12847  
Qy 2297 TCGTTTCTGTTAAATAGACCAATGCAACCTCCAGCTTCCAGTGAATAGATTAAGACTC 2356  
Db 12848 TCGTTTCTGTTAAATAGACCAATGCAACCTCCAGCTTCCAGTGAATAGATTAAGACTC 12907  
Qy 2357 AAACCATCCCTTGAAGCTGTCTTGGAAAGGGGCTTGGAGTGAACCCAGACTGTGGCTC 2416  
Db 12908 AAACCATCCCTTGAAGCTGTCTTGGAAAGGGGCTTGGAGTGAACCCAGACTGTGGCTC 12967  
Qy 2417 ATGGCTCATGTGATGTGAGTCAAGGCTCCCAACCTGTCAAGCAATTTGCTCCGTAGGA 2476  
Db 2417 ATGGCTCATGTGATGTGAGTCAAGGCTCCCAACCTGTCAAGCAATTTGCTCCGTAGGA 2476

Db 12968 ATGGCTCATGTGATGTGAGTCAAGGCTCCCAACCTGTCAAGCAATTTGCTCCGTAGGA 13027  
Qy 2477 CTTTGAATGGGTAGAGTATAGTAACTAAGCTCTGATCTGACACAAAGGCTTTGACTGGG 2536  
Db 13028 CTTTGAATGGGTAGAGTATAGTAACTAAGCTCTGATCTGACACAAAGGCTTTGACTGGG 13087  
Qy 2537 AGGCCAGGCTATAGAGTGTGCTCAGGCTTAAAGGGCTGGAGGCTGGGGGACAGTGTCTCAG 2596  
Db 13088 AGGCCAGGCTATAGAGTGTGCTCAGGCTTAAAGGGCTGGAGGCTGGGGGACAGTGTCTCAG 13147  
Qy 2597 ATTAGGCTTAACTAGGAATTTGACTGAGCTGAGAACAGAGTTAGGGGCCAAGCAGCA 2656  
Db 13148 ATTAGGCTTAACTAGGAATTTGACTGAGCTGAGAACAGAGTTAGGGGCCAAGCAGCA 13207  
Qy 2657 GGGTTGTGGGTCTACTCTCTTAGAGAGACCTTGAAGCTTACTTTTCACTTCTTAATGTTGTC 2716  
Db 13208 GGGTTGTGGGTCTACTCTCTTAGAGAGACCTTGAAGCTTACTTTTCACTTCTTAATGTTGTC 13267  
Qy 2717 TTGATGAGCTACCCCTCAAGGGGTTGGCTGCTAGTCTTAAGGGGTGAGACAAAGGACAGT 2776  
Db 13268 TTGATGAGCTACCCCTCAAGGGGTTGGCTGCTAGTCTTAAGGGGTGAGACAAAGGACAGT 13327  
Qy 2777 TTCAAGTCTGATCTTATCAAGTTCAATGCACTAACAATTGGGACCACTGCTGATCAATGCC 2836  
Db 13328 TTCAAGTCTGATCTTATCAAGTTCAATGCACTAACAATTGGGACCACTGCTGATCAATGCC 13387  
Qy 2837 AGGAGCTTAAAGAGTGTCTTAAACAGTTATCCAAACAATGTGATCCCAAGTTAACTTTC 2896  
Db 13388 AGGAGCTTAAAGAGTGTCTTAAACAGTTATCCAAACAATGTGATCCCAAGTTAACTTTC 13447  
Qy 2897 TCTTGTTCAGAGGAGGAGGATCTAAGTCTCCCTTCTCCTTCTCCCAAGTGTTC 2956  
Db 13448 TCTTGTTCAGAGGAGGAGGATCTAAGTCTCCCTTCTCCTTCTCCCAAGTGTTC 13507  
Qy 2957 TCTTGAAGGAAATCCTTAGCTTGTCTCCAGGGAATCCCCAGAAATGTTTGTTCAGTTC 3016  
Db 13508 TCTTGAAGGAAATCCTTAGCTTGTCTCCAGGGAATCCCCAGAAATGTTTGTTCAGTTC 13567  
Qy 3017 AGTTTAGGCTGTATTAAGGAATATCTTAGAGTGGGTATCTATACGAATAGAAATTTA 3076  
Db 13568 AGTTTAGGCTGTATTAAGGAATATCTTAGAGTGGGTATCTATACGAATAGAAATTTA 13627  
Qy 3077 TTGTTCACAATCTGAGGCTGGAATAATCAAGATCAAGGCTCCAGCAGTTCAGTGTCT 3136  
Db 13628 TTGTTCACAATCTGAGGCTGGAATAATCAAGATCAAGGCTCCAGCAGTTCAGTGTCT 13687  
Qy 3137 GCTGAGTGTCTTGTCTTCCAGAGTGTGCACTTTTGTCTGTGTCTCA 3186  
Db 13688 GCTGAGTGTCTTGTCTTCCAGAGTGTGCACTTTTGTCTGTGTCTCA 13737

RESULT 3  
US-10-125-237-19  
; Sequence 19, Application US/10125237  
; Publication No. US20030022329A1  
GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Zhou, Ping  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Zhang, Jie  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and  
FILE REFERENCE: 791CIP2ADiv  
CURRENT FILING DATE: US/10/125,237  
PRIOR APPLICATION NUMBER: 09/668,317



```

QY 2401 ACCGAGACTGGCTCATGGCCCTCATGTATCTGAGTACAGCCCTCCCACTGTACAGC 2460
DB 2401 ACCGAGACTGGCTCATGGCCCTCATGTATCTGAGTACAGCCCTCCCACTGTACAGC 2460
QY 2461 CATTTGCTCCGTAGAGACTTTGATGGGTAGAGTACAGTCAAGAGCTTGTACCTGACAC 2520
DB 2461 CATTTGCTCCGTAGAGACTTTGATGGGTAGAGTACAGTCAAGAGCTTGTACCTGACAC 2520
QY 2521 AAGGCTTTTACTGGAGGCGAGGCTATAGAGTGGCTCCAGCTTAAAGGCTGGAGCTG 2580
DB 2521 AAGGCTTTTACTGGAGGCGAGGCTATAGAGTGGCTCCAGCTTAAAGGCTGGAGCTG 2580
QY 2581 GGGGACAGTGTCTCAGATTAGGCTCTAAGTAAAGTGAAGTGAAGTGAAGAGAGT 2640
DB 2581 GGGGACAGTGTCTCAGATTAGGCTCTAAGTAAAGTGAAGTGAAGTGAAGAGAGT 2640
QY 2641 TAGGGGCGAAGCAGAGGCTGTGGGTCTATCTCTTAAAGACCTTGAAGCTTTACTTTT 2700
DB 2641 TAGGGGCGAAGCAGAGGCTGTGGGTCTATCTCTTAAAGACCTTGAAGCTTTACTTTT 2700
QY 2701 CATTCCTATAGGTCTTGGATAGGCTACCTCAAGGGGTGGCTGCTAGTCTTAAAGGGTG 2760
DB 2701 CATTCCTATAGGTCTTGGATAGGCTACCTCAAGGGGTGGCTGCTAGTCTTAAAGGGTG 2760
QY 2761 GAGACAAAGACAGAGTTCAGGTCTGCTCTTATCAAGTTCAGTCACTACCTTGGAGCC 2820
DB 2761 GAGACAAAGACAGAGTTCAGGTCTGCTCTTATCAAGTTCAGTCACTACCTTGGAGCC 2820
QY 2821 ACTGTGATCATGCGCAGGAGCTTAAAGGTGTCTAAAGATTATCAACATCTGTATAT 2880
DB 2821 ACTGTGATCATGCGCAGGAGCTTAAAGGTGTCTAAAGATTATCAACATCTGTATAT 2880
QY 2881 CCCAAGTTAACTTCTCTTGTGTTTGAAGGCGAGGAGTAAAGTCTCCCTTCTCTCT 2940
DB 2881 CCCAAGTTAACTTCTCTTGTGTTTGAAGGCGAGGAGTAAAGTCTCCCTTCTCTCT 2940
QY 2941 TTCCTCCAGTGTCTCTTTCAGAGGAACTCTTACCTGTCTCCAGGGAATCCAGAA 3000
DB 2941 TTCCTCCAGTGTCTCTTTCAGAGGAACTCTTACCTGTCTCCAGGGAATCCAGAA 3000
QY 3001 ATGTTTGTTCAGTCAAGTTTAAAGGCTGTAAAGAAATATCTTAAAGTGGTAACTAT 3060
DB 3001 ATGTTTGTTCAGTCAAGTTTAAAGGCTGTAAAGAAATATCTTAAAGTGGTAACTAT 3060
QY 3061 CAGCAATAGAAATTTATTTGTTCACATTTGAGAGGCTGAGAAATCCAAATCAAGCTCC 3120
DB 3061 CAGCAATAGAAATTTATTTGTTCACATTTGAGAGGCTGAGAAATCCAAATCAAGCTCC 3120
QY 3121 AGCAGGTCAGTGTCTGAGTGTGTTCTGCTTCAAGATGAGCACTTTTGTGCTG 3180
DB 3121 AGCAGGTCAGTGTCTGAGTGTGTTCTGCTTCAAGATGAGCACTTTTGTGCTG 3180
QY 3181 TTCTCA 3186
DB 3181 TTCTCA 3186

```

```

; TYPE: DNA
; ORGANISM: Human
US-09-900-448-3
Query Match 71.5%; Score 2279.2; DB 13; Length 13737;
Best Local Similarity 90.9%; Pred. No. 0;
Matches 2554; Conservative 0; Mismatches 3; Indels 253; Gaps 2;

QY 630 TTCAGTGGGAGCCACTACTGGGGTCTGACACACAGCCGGGATGGGATGCTGGCC 689
DB 10928 TCCACAGGAGCCACTACTGGGGTCTGACACACAGCCGGGATGGGATGCTGGCC 10987
QY 690 ATTGCTCATCAGTGGCCAGGAGTCTTCAAGAGTGAATGCTCTTTCTGGAGAA 749
DB 10988 ATTGCTCATCAGTGGCCAGGAGTCTTCAAGAGTGAATGCTCTTTCTGGAGAA 11047
QY 750 AAATCTATCTGG----- 762
DB 11048 AAATCTATCTGGTCCAGGTGTATTTGGGAGAGGCTTGAAGTGAAGACTGGACAAG 11107
QY 763 -----TCCAGGGCACCCAGTATATGCT 786
DB 11108 CATATCCAACTGTATTTATTTATTAACATCTTTGCTCCAGGACACCCAGTATATGCT 11167
QY 787 TCCGACAAAGGAGGCTATACCTTATAGGCGTTATCCGAAGCGGCTGAGAAAG 846
DB 11168 TCCGACAAAGGAGGCTATACCTTATAGGCGTTATCCGAAGCGGCTGAGAAAG 11227
QY 847 TGGGAGCCCTCATAGGATTTATCTGAGCTGTGTGATGCGGCTTTATCTGCTGGGT 906
DB 11228 TGGGAGCCCTCATAGGATTTATCTGAGCTGTGTGATGCGGCTTTATCTGCTGGGT 11287
QY 907 CTTTCGGGCTCCATATCATAG----- 927
DB 11288 CTTTCGGGCTCCATATCATAGGAGGAGGCTTCTGGGTCTTAAAGGAGACTTGT 11347
QY 928 ----- 927
DB 11348 CTGTACTGTCTGTGGCATAGATCCCAACAGAGGATGAGAAAGGCTTAGATCC 11407
QY 928 ----- 927
DB 11408 CCAGGACATGAGAGGCTTAGATCCCAATGACATGAGAGCCATGATGTTGG 11467
QY 928 -----CAGACGGGCGCTGTGTGGCTGACCTGAACTGAGAGCCCAAGCCAC 976
DB 11468 TGCCTTCTCCAGAGAGGCGGCTGTGTGGCTGACCTGAACTGAGAGCCCAAGCCAC 11527
QY 977 GTGACAGAGCTTCTTGGGCCCATGAGAGGTTAGACGAGGCTGTGTATGAGAAAGTC 1036
DB 11528 GTGACAGAGCTTCTTGGGCCCATGAGAGGTTAGACGAGGCTGTGTATGAGAAAGTC 11587
QY 1037 CCTTGGCCCTTAACTCATGTTCGCCAATGTCGCCGCTTGTAACTTCATCATGTGCCAA 1096
DB 11588 CCTTGGCCCTTAACTCATGTTCGCCAATGTCGCCGCTTGTAACTTCATCATGTGCCAA 11647
QY 1097 TTTGTACTGCTACAGTGAATGTGAGAAATGATGACGCAAGCCCTTCCCAACCCCA 1156
DB 11648 TTTGTACTGCTACAGTGAATGTGAGAAATGATGACGCAAGCCCTTCCCAACCCCA 11707
QY 1157 GAATGACAGAGTCTCTGGGCTGACACTCAGAGGAGCTTCTGACATGAGTGGCT 1216
DB 11708 GAATGACAGAGTCTCTGGGCTGACACTCAGAGGAGCTTCTGACATGAGTGGCT 11767
QY 1217 GGGCCCACTCTAGTCTCTCATTAATAAGACAGATGCTTGTGCTTCTGACAGAGG 1276
DB 11768 GGGCCCACTCTAGTCTCTCATTAATAAGACAGATGCTTGTGCTTCTGACAGAGG 11827
QY 1277 GCTTCTGACATGAGTCTGGCTGGCCCACTCCCAAGTTTCTCATTAATAAGACAGAT 1336
DB 11828 GCTTCTGACATGAGTCTGGCTGGCCCACTCCCAAGTTTCTCATTAATAAGACAGAT 11887
QY 1337 TGCTTCTCACTTGAATCAAGGAGCCTTGTGTGTAACATCTTCTTGAAGTTGA 1396

```

Db 181 TTGATGCTACCAACCTTGATGACATGTAAGAAACATGCTGTTTTTAAAGGGAGTTTGCT 240  
Qy 241 GGAAGAGTCAAAATGGGACCGGAGTTAATCTCAGAGAGATGGAAGAAATTTCCCAAGCC 300  
Db 241 GGAAGAGTCAAAATGGGACCGGAGTTAATCTCAGAGAGATGGAAGAAATTTCCCAAGCC 300  
Qy 301 CTGTGATGCTGATTCGGTCAAGGTCAACAAGTGTCTTTCATCAACAGGGGACAAAG 360  
Db 301 CTGTGATGCTGATTCGGTCAAGGTCAACAAGTGTCTTTCATCAACAGGGGACAAAG 360  
Qy 361 TCTGGGTATACCTCTCTGAAAAGAGAAAGATACCCAAAGTTGCTCCAAAGATGAAT 420  
Db 361 TCTGGGTATACCTCTCTGAAAAGAGAAAGATACCCAAAGTTGCTCCAAAGATGAAT 420  
Qy 421 TTCTGGGATATCCATCCCTGAAAAGAGAAAGATACCCAAAGTTGCTCCAAAGATGAAT 420  
Db 421 TTCTGGGATATCCATCCCTGAAAAGAGAAAGATACCCAAAGTTGCTCCAAAGATGAAT 420  
Qy 481 CTGAAGGCGTCTCTTCTTCAAGGCGATGGAACAAGAAATGGGACTGGGCAATGGGAACA 540  
Db 481 CTGAAGGCGTCTCTTCTTCAAGGCGATGGAACAAGAAATGGGACTGGGCAATGGGAACA 540  
Qy 541 GTACCCACATGAGCCCTGAGTATATGCGTGTAGCCCAATCTAATGTTGTCTGCACTGA 600  
Db 541 GTACCCACATGAGCCCTGAGTATATGCGTGTAGCCCAATCTAATGTTGTCTGCACTGA 600  
Qy 601 CGTCTGCAACATGCTGCACTATGCTTCAAGTGGGACCCCACTACTGCGCTGTGACA 660  
Db 601 CGTCTGCAACATGCTGCACTATGCTTCAAGTGGGACCCCACTACTGCGCTGTGACA 660  
Qy 661 CCAGCGGGAGTGGCTGGCATAGCTGGCCCATGCTCATCAATGGGCCCAAGGCTCTTCAAG 720  
Db 661 CCAGCGGGAGTGGCTGGCATAGCTGGCCCATGCTCATCAATGGGCCCAAGGCTCTTCAAG 720  
Qy 721 CAGTGTAGTGTGCTTTCTCTGGAAGAAAACCTATCTGTCTCAAGGACCCAGGATAT 780  
Db 721 CAGTGTAGTGTGCTTTCTCTGGAAGAAAACCTATCTGTCTCAAGGACCCAGGATAT 780  
Qy 781 ATGTCTTCTGCAAAAGGAGGCTATACCTAGTAAGCGGTATACCAACCGGCTGGAGA 840  
Db 781 ATGTCTTCTGCAAAAGGAGGCTATACCTAGTAAGCGGTATACCAACCGGCTGGAGA 840  
Qy 841 AGGAAGCGGGAGCCCTCATGGGATTAATCTGGACTCTGGAATGGGCGCTTATCTGCC 900  
Db 841 AGGAAGCGGGAGCCCTCATGGGATTAATCTGGACTCTGGAATGGGCGCTTATCTGCC 900  
Qy 901 CTGGGCTTCTCGGCTCATATCATGSCAAGAGCGGCGCTGTGGTGGCTGGACTGAAGT 960  
Db 901 CTGGGCTTCTCGGCTCATATCATGSCAAGAGCGGCGCTGTGGTGGCTGGACTGAAGT 960  
Qy 961 CAGGAGCCCAAGCGTGAACAAGAGTCTTGGGCCCATGAGAAAGTGAAGGAGGCT 1020  
Db 961 CAGGAGCCCAAGCGTGAACAAGAGTCTTGGGCCCATGAGAAAGTGAAGGAGGCT 1020  
Qy 1021 TGGTATGGAAGAAAGTCCCTTGGCCCTTAATCAATGTTCCGGCAATGGGCGGCTTGAAC 1080  
Db 1021 TGGTATGGAAGAAAGTCCCTTGGCCCTTAATCAATGTTCCGGCAATGGGCGGCTTGAAC 1080  
Qy 1081 TCAATCAGGTGCCAATTTGTACTGCTACAGTATGTGAGAAACTGAATGACCCAAAG 1140  
Db 1081 TCAATCAGGTGCCAATTTGTACTGCTACAGTATGTGAGAAACTGAATGACCCAAAG 1140  
Qy 1141 CCGTTTCGCAACCCCAAGATGTGACAGTCTCTGGGCTGCACTCACTGAGGGGCTTCT 1200  
Db 1141 CCGTTTCGCAACCCCAAGATGTGACAGTCTCTGGGCTGCACTCACTGAGGGGCTTCT 1200  
Qy 1201 GACATGATGTGGCTGGGCCCACTCTCTAGTTCTCATATATTAAGCAATTTGCTTCT 1260  
Db 1201 GACATGATGTGGCTGGGCCCACTCTCTAGTTCTCATATATTAAGCAATTTGCTTCT 1260  
Qy 1261 CGGCTTCTCACTGAGGGGCTTCTGACATGAGTCTGGGCTGGGCCCACTCTCCCAAGTTCT 1320  
Db 1261 CGGCTTCTCACTGAGGGGCTTCTGACATGAGTCTGGGCTGGGCCCACTCTCCCAAGTTCT 1320

Qy 1321 CATATTAAGAGAGATGCTTCTTCACTTGAATCAAGGGGACCTTGGTGGTGAACAATCT 1380  
Db 1321 CATATTAAGAGAGATGCTTCTTCTTCACTTGAATCAAGGGGACCTTGGTGGTGAACAATCT 1380  
Qy 1381 TCTTTCTTGAAGTGAAGAAAGTTAGCACTTCTCTTGAAGGAGTGTCAAGCTCAACAAGGC 1440  
Db 1381 TCTTTCTTGAAGTGAAGAAAGTTAGCACTTCTCTTGAAGGAGTGTCAAGCTCAACAAGGC 1440  
Qy 1441 TGTGAGAAAACAAGGAGGAGCACTAAGGGCAAACTATCTCTGCGAGATGATTTCTT 1500  
Db 1441 TGTGAGAAAACAAGGAGGAGCACTAAGGGCAAACTATCTCTGCGAGATGATTTCTT 1500  
Qy 1501 AGGTCCAGATCATTAACCTTGTCTTGAAGCACTATCTAACAATAGTGGGGGAAAAGAA 1560  
Db 1501 AGGTCCAGATCATTAACCTTGTCTTGAAGCACTATCTAACAATAGTGGGGGAAAAGAA 1560  
Qy 1561 CCAGAGTCCGAAGAGGAACAGCTGAGTTATACAGCAATGAAGAGGTGAGAGTCACTC 1620  
Db 1561 CCAGAGTCCGAAGAGGAACAGCTGAGTTATACAGCAATGAAGAGGTGAGAGTCACTC 1620  
Qy 1621 TGAATCACTTCTGTGATGATGGCCCAACCCAGCCGCAAGGCATGAGAAACAAGGCG 1680  
Db 1621 TGAATCACTTCTGTGATGATGGCCCAACCCAGCCGCAAGGCATGAGAAACAAGGCG 1680  
Qy 1681 CTGGGGCAACTATGCACTGTGCAAAAGAGATGGCTCAAGTTGTGGGCTTGAAGAGTCCA 1740  
Db 1681 CTGGGGCAACTATGCACTGTGCAAAAGAGATGGCTCAAGTTGTGGGCTTGAAGAGTCCA 1740  
Qy 1741 TCTGGGGACCTCAAAATTAATGCTTGGTGAATTCAGTAACAACAATCACTGAGCTTGT 1800  
Db 1741 TCTGGGGACCTCAAAATTAATGCTTGGTGAATTCAGTAACAACAATCACTGAGCTTGT 1800  
Qy 1801 TGCATGAGTTAAGCATGACAAGTGAATGAAGTTGAAGTGGGAAAACAAGAAATCACCA 1860  
Db 1801 TGCATGAGTTAAGCATGACAAGTGAATGAAGTTGAAGTGGGAAAACAAGAAATCACCA 1860  
Qy 1861 GCTGTGTGCAAGGCAACTGTGAGAGAGAGAAAGAAATGAATGATGACCAATGAGGAC 1920  
Db 1861 GCTGTGTGCAAGGCAACTGTGAGAGAGAGAAAGAAATGAATGATGACCAATGAGGAC 1920  
Qy 1921 ATTGCAAGAACAGTCCCTGGGAGTCTCTGAGAGCTCAAGGAGCTTGTGGCACAGA 1980  
Db 1921 ATTGCAAGAACAGTCCCTGGGAGTCTCTGAGAGCTCAAGGAGCTTGTGGCACAGA 1980  
Qy 1981 GGAATCGGCTACCAATTAAGCTCTGAGGATCTGACACATCTAGACCAAGCAATGTCA 2040  
Db 1981 GGAATCGGCTACCAATTAAGCTCTGAGGATCTGACACATCTAGACCAAGCAATGTCA 2040  
Qy 2041 CTGGCAAGAGGTTGCAAGTGTGTTATTTCTGTGATCAATACTGTGAAGGCTTTGGG 2100  
Db 2041 CTGGCAAGAGGTTGCAAGTGTGTTATTTCTGTGATCAATACTGTGAAGGCTTTGGG 2100  
Qy 2101 TTCCAAATTTGCTGACAGCTGTTTAACTGGGAAATGGGCTAGACTATAGTATGT 2160  
Db 2101 TTCCAAATTTGCTGACAGCTGTTTAACTGGGAAATGGGCTAGACTATAGTATGT 2160  
Qy 2161 CTGAGCAAGGCGCTATCTTCTCACTGCTTTAACAACCAAGCTGAGGTTGAGAGCTGCT 2220  
Db 2161 CTGAGCAAGGCGCTATCTTCTCACTGCTTTAACAACCAAGCTGAGGTTGAGAGCTGCT 2220  
Qy 2221 TGTTCAGGCTCAAAAATAAGCTGATTTCCAGCAGAGAGGCGCTTATCTGAGCTTGG 2280  
Db 2221 TGTTCAGGCTCAAAAATAAGCTGATTTCCAGCAGAGAGGCGCTTATCTGAGCTTGG 2280  
Qy 2281 TGTCTTGGCTCAATTTCTTCTCTGTAATAATGACACATGCAACCCACTTCCAGTGA 2340  
Db 2281 TGTCTTGGCTCAATTTCTTCTCTGTAATAATGACACATGCAACCCACTTCCAGTGA 2340  
Qy 2341 CAATGAATATGACTCAAAACCAATCCCTTGAACCTGCTTGGGAAGGGGCTTGGAGCTAG 2400  
Db 2341 CAATGAATATGACTCAAAACCAATCCCTTGAACCTGCTTGGGAAGGGGCTTGGAGCTAG 2400

Result No.	Score	Query Match	Length	DB	ID	Description
1	3186	100.0	3186	13	US-09-900-448-1	Sequence 1, App11
2	2279.2	71.5	13757	13	US-09-900-448-3	Sequence 3, App11
3	1129.8	35.5	1631	15	US-10-125-237-19	Sequence 19, App1
4	1129.8	35.5	1631	15	US-10-105-891-19	Sequence 19, App1
5	576.8	18.1	1516	13	US-10-316-253-39	Sequence 39, App1
6	576.8	18.1	1516	15	US-10-175-523-84	Sequence 84, App1
7	427	13.2	488	11	US-09-918-995-0828	Sequence 30828, App1
8	421	13.2	488	11	US-09-918-995-8595	Sequence 8595, App1
9	417	13.1	488	11	US-09-918-995-11395	Sequence 11395, App1
10	374.2	11.7	492	11	US-09-918-995-10754	Sequence 10754, App1
11	345.4	10.8	491	11	US-09-918-995-32820	Sequence 32820, App1
12	313.4	9.8	504	10	US-09-880-107-1166	Sequence 1166, App1
13	298.4	9.4	420	10	US-09-960-352-10323	Sequence 10323, App1
14	280	8.8	408	10	US-09-960-352-7040	Sequence 7040, App1
15	278	8.7	384	11	US-09-918-995-7738	Sequence 7738, App1

16	269.4	8.7	473	11	US-09-618-995-32181	Sequence 53181, Ap
17	269.2	8.4	374	10	US-09-660-352-5368	Sequence 5368, Ap
18	266.8	8.4	383	10	US-09-660-352-6886	Sequence 6846, Ap
19	261.4	8.3	437	10	US-09-660-352-380	Sequence 380, Ap
20	261.8	8.2	405	10	US-09-660-352-445	Sequence 445, Ap
21	260.2	8.1	425	10	US-09-660-352-3702	Sequence 3702, Ap
22	259	8.2	425	10	US-09-660-352-9378	Sequence 9378, Ap
23	257.4	8.1	397	10	US-09-660-352-12936	Sequence 12936, Ap
24	256	8.0	387	10	US-09-660-352-11316	Sequence 11316, Ap
25	256	8.0	422	10	US-09-660-352-7056	Sequence 7056, Ap
26	255.2	8.0	418	10	US-09-660-352-13395	Sequence 13395, Ap
27	255	8.0	373	10	US-09-660-352-2026	Sequence 2026, Ap
28	254.4	8.0	412	10	US-09-660-352-12507	Sequence 12507, Ap
29	251.4	7.9	433	10	US-09-660-352-13402	Sequence 13402, Ap
30	251.2	7.9	404	10	US-09-660-352-7514	Sequence 7514, Ap
31	250.2	7.9	412	10	US-09-660-352-8662	Sequence 8662, Ap
32	250.2	7.9	417	10	US-09-660-352-14176	Sequence 14176, Ap
33	250.2	7.9	431	10	US-09-660-352-7130	Sequence 7130, Ap
34	249.8	7.8	399	10	US-09-660-352-457	Sequence 457, Ap
35	249.4	7.8	407	10	US-09-660-352-136	Sequence 136, Ap
36	249.4	7.8	410	10	US-09-660-352-6107	Sequence 6107, Ap
37	249.4	7.8	411	10	US-09-660-352-5619	Sequence 5619, Ap
38	249.2	7.8	366	10	US-09-660-352-4209	Sequence 4209, Ap
39	248.6	7.8	407	10	US-09-660-352-5920	Sequence 5920, Ap
40	248.2	7.8	405	10	US-09-660-352-2956	Sequence 2956, Ap
41	244.6	7.7	407	10	US-09-660-352-13635	Sequence 13635, Ap
42	244.6	7.7	408	10	US-09-660-352-7025	Sequence 7025, Ap
43	243.6	7.6	392	10	US-09-660-352-911789	Sequence 1789, Ap
44	243.2	7.6	364	10	US-09-660-352-9132	Sequence 9132, Ap
45	239.4	7.5	397	10	US-09-660-352-4643	Sequence 4643, Ap

## ALIGNMENTS

```

RESULT 1
US-09-900-448-1
; Sequence 1, Application US/09900448
; Publication No. US20030220488A1
; GENERAL INFORMATION:
; APPLICANT: CECCARDI, Toni et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO01272
; CURRENT APPLICATION NUMBER: US/09/900,448
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3186
; TYPE: DNA
; ORGANISM: Human
; US-09-900-448-1

```

	Query Match	Similarity	100.0%	Score 3186	DB 13	Length 3186	
	Best Local	Similarity	100.0%	Pred. No. 0	Mismatches 0	Indels 0	Gaps 0
	Matches 3186	Conservative	0				
Qy	1	CTCTGAGCTCAGCATGCTTAGGTTACTGGAGACACCGTTTGCACTGGGTTTGTGAGCC	60				
Db	1	CTCTGAGGCTCAGCATGCTTAGGTTACTGGAGACACCGTTTGCACTGGGTTTGTGAGCC	60				
Qy	61	TATGTGCTCTTGAGCCATTGGCACCCTCTTCTCCGCACTAGTGCCCATGGGAATGTTG	120				
Db	61	TATGTGCTCTTGAGCCATTGGCACCCTCTTCTCCGCACTAGTGCCCATGGGAATGTTG	120				
Qy	121	CTGAAAGCGAGACCAAGCCAGACCCAGCTGAAACGCTGTCAGATGGCTTGAGACT	180				
Db	121	CTGAAAGCGAGACCAAGCCAGACCCAGCTGAAACGCTGTCAGATGGCTTGAGACT	180				
Qy	181	TTGATGCTACACCCCTGGATGACATGGAACCATGCTGTTTTTTTAAAGGCGAGTTTGCT	240				



```
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 197
LENGTH: 1282
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (675)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1195)
OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-197
```

Query Match 2.2%; Score 69.2; DB 4; Length 1282;

Best Local Similarity 73.0%; Pred. No. 5.4e-12; Matches 89; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 3003 GGTTCCTTCAGTCAGTTAGGCTGCTATAGAGAAATCTTAGAGTGGTAATCTATCA 3062
DB 505 GGTATGTAATCTTCATTTTGTCTGCTATACAGAAATACAGAGACTGGTAATTTGTAA 446
QY 3063 GCATATGAAATTTATTTGTTTCAATTTCTGAGGCTGGAAATCAAGATCAAGGCTCCAG 3122
DB 445 GGAACGAATATTTATTTCTCAGATTTCTGAGGATGGAGATCAAGATCAAGAGCTGG 386
QY 3123 CA 3124
DB 385 CA 384
```

## RESULT 14

```
US-09-146-053-3/c
Sequence 3, Application US/09146053A
Patent No. 639349
GENERAL INFORMATION:
APPLICANT: Ryan, James W.
APPLICANT: Sprinkle, Terry Joe Curtis
APPLICANT: Venema, Richard C.
TITLE OF INVENTION: Human Aminopeptidase P Gene
FILE REFERENCE: MCG103
CURRENT APPLICATION NUMBER: US/09/146,053A
CURRENT FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/057,854
EARLIER FILING DATE: 1997-09-02
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 50000
TYPE: DNA
ORGANISM: Homo sapiens
US-09-146-053-3
```

Query Match 2.2%; Score 69; DB 4; Length 50000;  
Best Local Similarity 72.4%; Pred. No. 8.6e-11;  
Matches 131; Conservative 0; Mismatches 45; Indels 5; Gaps 3;

```
QY 3007 TGTTCAGTCAGTTAGGCTGCTATAGAGAAATCTTAGAGTGGTAATCTATCAGCA 3066
DB 37631 TGTTCAGTCAGTTAGGCTGCTATAGAGAAATCTTAGAGTGGTAATCTATCAGCA 37575
QY 3067 TGAATTTATTTATTTTCAATTTCTGAGGCTGGAAATCAAGATCAAGGCTCCAGCAG 3126
DB 37574 TGAATTTATTTATTTTCAATTTCTGAGGCTGGAAATCAAGATCAAGGCTCCAGCAG 37515
QY 3127 TTCAG-TGTTCGCTGAGTCTGT-TTCGCTTCAAGATGAGCACTTTTGTCTGTCT 3184
DB 37514 TTCAGTGTCTGTGAGGCTGAGGCTCTTCATCAAGATGAGCACTTTGTCTGTCT 37455
QY 3185 C 3185
DB 37454 C 37454
```

## RESULT 15

```
US-09-877-177A-10/c
Sequence 10, Application US/09877177A
Patent No. 6582919
GENERAL INFORMATION:
APPLICANT: K. Danenberg
TITLE OF INVENTION: Method of determining Epidermal Growth
TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression
TITLE OF INVENTION: and Correlation of Levels Thereof with Survival
FILE REFERENCE: 11220/120
CURRENT APPLICATION NUMBER: US/09/877,177A
CURRENT FILING DATE: 2001-06-11
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 107496
LENGTH: 197496
TYPE: DNA
ORGANISM: Homo sapiens
US-09-877-177A-10
```

Query Match 2.1%; Score 68.4; DB 4; Length 197496;

Best Local Similarity 67.6%; Pred. No. 3.6e-10; Matches 96; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
QY 3004 GTTTCCTTCAGTCAGTTAGGCTGCTATAGAGAAATCTTAGAGTGGTAATCTATCAG 3063
DB 3500 GTATGCTTCATCTTTTGGGCTTCTGTACAAATATACATTAATGATTTACTGTAA 3441
QY 3064 CAATAGAAATTTATTTGTTTCAATTTCTGAGGCTGGAAATCAAGATCAAGGCTCCAGC 3123
DB 3440 CAACAGAAATTTATTTCCATAGTTCTGAGGCTGGAAATCAAGATCAAGGATTTTGC 3381
QY 3124 AGTTTCAGTCTGCTGAGTGC 3145
DB 3380 AGATTTGATGTCTAATAGAGC 3359
```

Search completed: December 16, 2003, 05:46:23  
Job time: 141.986 secs

;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090431  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090435  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090444  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090472  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090535  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090540  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090542  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090676  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090678  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090690  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090694  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090695  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090696  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090862  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091478  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091544  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091626  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091633  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/092182  
;; PRIOR FILING DATE: 1998-07-09

Query Match 2.2%; Score 69.8; DB 4; Length 762;

Best Local Similarity 71.3%; Pred. No. 2,4e-12; Matches 92; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 3058 TATCAGAAATGAGATTATTTGTTCAAAATTCGAGGCTGGAAATCCAAATCAAGGC 3117  
DB 756 TATTAACAGCAAAATATTATTTCTTACAGTTCTAGAGCTGGGAAGTCCAAAGTCAAGGC 697  
QY 3118 TCCAGCAGTTGCTGCTGCTGAGTGTCTTGTCTGCTGGAAGATGCGACCTTTTGT 3177  
DB 696 ACCAGCAGATTCCAGATCTGTGTAGGACCAAGTTCTCTTCTGACAGTGCCTTACGCT 637  
QY 3178 GTGTTCTCA 3186  
DB 636 GTGATCACA 628

RESULT 13  
US-09-205-258-197/c

;; Sequence 197, Application US/09205258  
;; Patent No. 6525174  
;; GENERAL INFORMATION:  
;; APPLICANT: Young et al.  
;; TITLE OF INVENTION: 207 Human Secreted Proteins  
;; FILE REFERENCE: P2007P1  
;; CURRENT FILING DATE: 1998-12-04  
;; EARLIER APPLICATION NUMBER: PCT/US98/11422  
;; EARLIER FILING DATE: 1998-06-04  
;; EARLIER APPLICATION NUMBER: 60/048,885  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/049,375  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,881  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,880  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,896  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/049,020  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,876  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,895  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,884  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,894  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,971  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,964  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,882  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,899  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,893  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,900  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,901  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,892  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,915  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/049,019  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,970  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,972  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,916  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/049,373  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,875  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/049,374  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,917  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,949  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,974  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,883  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,897  
;; EARLIER FILING DATE: 1997-06-06  
;; EARLIER APPLICATION NUMBER: 60/048,898

Patent No. 6478825  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas P.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Thomas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2730P1C3  
CURRENT APPLICATION NUMBER: US/09/996,243  
CURRENT FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/075945  
PRIOR FILING DATE: 1998-02-25  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088021  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088030  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088326  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088202  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088212  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088734  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088742  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089440  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429



Db 900 ACAGAAATGATGCTCAGAGCTGGAGGCTGGAAGATTCAGATCAAGCCGCGAG 959  
QY 3126 GTTCAGTCTGCTGAGATGCTGTTCT 3152  
Db 960 TTTCGGTGTGTGGAGAGCTTTGCT 986

## RESULT 10

US-08-611-587-1/c  
Sequence 1, Application US/08611587  
Patent No. 6150091

GENERAL INFORMATION:  
APPLICANT: PANDOLFO, MASSIMO  
APPLICANT: MONTERMINI, LAURA  
APPLICANT: MOLTO, MARIA D.  
APPLICANT: Koenig, Michael  
APPLICANT: Campuzano, Victoria  
TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.  
ZIP: 77010

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,587  
FILING DATE: 03-MAR-1996

## CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:  
NAME: Brashears-Macatee, Sarah J.  
REGISTRATION NUMBER: 38,087  
REFERENCE/DOCKET NUMBER: D-5901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5620  
TELEFAX: 713-651-5246

TELEX: 76-2829  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8353 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 9q13  
UNITS: bp

US-08-611-587-1

Query Match 2.5%; Score 78.4; DB 3; Length 8353;  
Best Local Similarity 66.8%; Pred. No. 1.9e-14;  
Matches 127; Conservative 0; Mismatches 61; Indels 2; Gaps 1;

QY 2996 CAGAAATGTTGTTTTCAGTCAAGTTCCTAATAAGAAATATCTTAGAGTGGTAA 3055  
Db 8244 CAAATATTTTGTATTTGCTTAGCTGGAGTGTATAAACAATATCCAGCACTGGGTGG 8185  
QY 3056 TCTATCGCAATGAGATTTATGTTTCAATTTCTGAGAGCTGGAATATCCAGATCAAG 3115  
Db 8184 CTTATTAACAATGAAATGTTTCTTAGAGTCTCTGAGGAGTGGGAATCAAGATCAAT 8125

QY 3116 GCTCCAGCAGATTCAGTGTCTGAGTGTCTT--CTGCTTGAAGATGAGCCTTTT 3173  
Db 8124 GCACCAACAGCTTTGGTGTCTGTGAGGGCAGTTTCTGTGTTCAAGATGAGCCTTGT 8065  
QY 3174 TGCTGTGTTT 3183  
Db 8064 CACTGGGCTC 8055

## RESULT 11

US-09-009-913-1/c  
Sequence 1, Application US/09009913  
Patent No. 6087485

GENERAL INFORMATION:  
APPLICANT: AYS Pharmaceuticals, Inc.  
TITLE OF INVENTION: Asthma Related Genes  
NUMBER OF SEQUENCES: 339  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bozicevic & Reed, LLP  
STREET: 285 Hamilton Ave, Suite 200  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94301

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,913  
FILING DATE: 21-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: SEQ-4P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3231  
TELEFAX: 650-327-3231

TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 72928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA

US-09-009-913-1

Query Match 2.4%; Score 77.6; DB 3; Length 72928;  
Best Local Similarity 72.2%; Pred. No. 1.6e-13;  
Matches 130; Conservative 0; Mismatches 44; Indels 6; Gaps 2;

QY 3009 TTTCAGTCAAGTTCAGTCTGTATTAAGAAATATCTTAGAGTGGTAAATATCAACAATA 3068  
Db 36263 TCTTAGTCAATTTACCTGCAACAAAGATGCTAGACTAGG---CTACAGCAACG 36208  
QY 3069 GGAATTTATTTGTTTCAAAATTCGAGAGCTGGAATAATCCAGATCAAGGCTCCAGAGGTT 3128  
Db 36207 GAAATTTATTTATTTCAAAATTTCTAGAGCTGGGAAGTCCAGAGCAAGGCAAGGAGATT 36148  
QY 3129 CAGTGTCTGTAGTGTG--CTGTGTTCTGCTTGAAGATGAGCACTTTTGTGTGTTCTCA 3186  
Db 36147 CAGCGTGTGTAGGGCCCACTTCTTAGTTCATAGATGCAATCTTCTCACCGTGAACCTCA 36088

## RESULT 12

US-09-996-243-344/c  
Sequence 344, Application US/09996243

EARLIER FILING DATE: 1997-09-02  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 50000  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-146-053-4

Query Match  
Best Local Similarity 68.7%; Score 80.4; DB 4; Length 50000;  
Pred. No. 1.5e-14;  
Matches 125; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

3005 TTGTTGACGTCAGTTAGGCTGCTGATTAAGAGATATCTTAAGTGGGTATCTATCAGC 3064  
37896 TTATCTTCACTTCATTGGGCTGCTGATTAAGAGAGCTGCAAGCTGTTGGCTTATTAAC 37955

3065 AATAGGAATTATTTGTTCAATCTTGAGAGCTGGAATAATCCAAGATCAAGGCTCCAGCA 3124  
37956 AGTAGACATTATTTCTTACAGTTCTGAGAGCTGGGAAGTGAAGATCAAGGACCTAGCA 38015

3125 GGTTCAGTGTCTGCTGAGTCTGTTGCTTGTGGAAGATGGACCTTTTGTGTGTCTT 3184  
38016 GATTGGTGTGCAACAGGCGCCAGT-TCCTTGTGATGCAACCTCTAGCTGTATCCT 38074

3185 CA 3186  
38075 CA 38076

RESULT 8  
US-09-671-317-212  
Sequence 212, Application US/09671317  
Patent No. 6528260  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIOMOLECULAR MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM  
FILE REFERENCE: 62, US3, CIP  
CURRENT APPLICATION NUMBER: US/09/671,317  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US 09/536,178  
PRIOR FILING DATE: 2000-03-23  
PRIOR APPLICATION NUMBER: PCT/IB00/00403  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: US 60/126,269  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/131,961  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 977  
SOFTWARE: Patent.pm  
SEQ ID NO 212  
LENGTH: 1001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 12-500-220 : polymorphic base A or G  
NAME/KEY: misc binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-500-220. misc2, potential complement  
NAME/KEY: primer bind  
LOCATION: 283..303  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer bind  
LOCATION: 711..731  
OTHER INFORMATION: downstream amplification primer, complement

NAME/KEY: misc binding  
LOCATION: 489..513  
OTHER INFORMATION: 12-500-220 potential probe  
US-09-671-317-212

Query Match  
Best Local Similarity 75.5%; Score 79; DB 4; Length 1001;  
Pred. No. 2.7e-15;  
Matches 111; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

3007 TGTTCAGTCAGTTAGGCTGCTGATTAAGA-GAATATCTTAGAGTGGTATATCATCAGCA 3065  
837 TGTCTGGTTGTTTGGGCTGCTGTAACAAATACCTTACAGTGGGTGCTTATTAACA 896

3066 ATAGGAATTATTTGTTCAATCTTGAGAGCTGGAATAATCCAAGATCAAGGCTCCAGCAG 3125  
897 ACGGAATGATTGCTCAGAGTTCTGAGAGCTGGGAAGTTCAGATCAAGTACCGGCGAG 956

3126 GTTCACTGTCTGCTGAGTCTGTTCT 3152  
957 TTTCGGTGTGTGAGAGCTTTTCT 983

RESULT 9  
US-09-671-317-458  
Sequence 458, Application US/09671317  
Patent No. 6528260  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIOMOLECULAR MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM  
FILE REFERENCE: 62, US3, CIP  
CURRENT APPLICATION NUMBER: US/09/671,317  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US 09/536,178  
PRIOR FILING DATE: 2000-03-23  
PRIOR APPLICATION NUMBER: PCT/IB00/00403  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: US 60/126,269  
PRIOR FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/131,961  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 977  
SOFTWARE: Patent.pm  
SEQ ID NO 458  
LENGTH: 1001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 12-500-217 : insertion CAATA  
NAME/KEY: misc binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-500-217. misc1, potential  
NAME/KEY: primer bind  
LOCATION: 286..306  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer bind  
LOCATION: 714..734  
OTHER INFORMATION: downstream amplification primer, complement  
US-09-671-317-458

Query Match  
Best Local Similarity 75.5%; Score 79; DB 4; Length 1001;  
Pred. No. 2.7e-15;  
Matches 111; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

3007 TGTTCAGTCAGTTAGGCTGCTGATTAAGA-GAATATCTTAGAGTGGTATATCATCAGCA 3065  
840 TGTCTGGTTGTTTGGGCTGCTGTAACAAATACCTTACAGTGGGTGCTTATTAACA 899

3066 ATAGGAATTATTTGTTCAATCTTGAGAGCTGGAATAATCCAAGATCAAGGCTCCAGCAG 3125

LOCATION: 68697  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68718  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68733  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68739  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 69785  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 79134  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 79198  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 86336  
OTHER INFORMATION: unknown  
OTHER INFORMATION: unknown  
US-09-791-211-3

Query Match 2.7%; Score 86; DB 4; Length 87543;  
Best Local Similarity 73.3%; Pred. No. 3.1e-16;  
Matches 151; Conservative 0; Mismatches 50; Indels 5; Gaps 3;

QY 2983 TCCAGGAACTCCAGAAATGTTGTTTCAGTCAGTTAGGCTGCT--ATAAGAAAT 3039  
DB 66969 TACTAGTAGTCGACAGACTGATTTGCTAGTCTGTTGCTGTTTCATACAAAT 68910  
QY 3040 ATCTTAGAGTGGTAACTATTCAGCATAGCAATTTATTTGTCACATTCGAGGCTGG 3099  
DB 68909 GCCACAGACTGGGCAATTTCAACAACAGATGTTATTTCTATAGTTCTGGAGGCTGG 68850  
QY 3100 AAAATCCAGATCAGGCTCCAGAGGTTCACTGCTGCTGAGTCTTGT-TCTGCTTCG 3158  
DB 68849 -AGGTCCAGATCAAGGCGCCAGTAGGTTCACTGCTGAGGAGGCTGCTCTGCTTCC 68791  
QY 3159 AAGATGGACCTTTTGGCTGCTGTTCT 3184  
DB 68790 AAAATGAGGCTTGTCTGCTGTTCT 68765

RESULT 5  
US-09-804-471A-3

Sequence 3, Application US/09804471A  
Patent No. 6479269  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
FILE REFERENCE: CL001164  
CURRENT APPLICATION NUMBER: US/09/804,471A  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 174493  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(174493)  
OTHER INFORMATION: n = A,T,C or G  
US-09-804-471A-3

Query Match 2.6%; Score 82.6; DB 4; Length 174493;  
Best Local Similarity 72.2%; Pred. No. 6.8e-15;

Matches 135; Conservative 0; Mismatches 49; Indels 3; Gaps 2;  
QY 3000 AATGTTTTCAGTCAGTTAGGCTGCTATTAAGAAATATCTTAGAGTGGTATCTA 3059  
DB 14097 AACGTGATCTGAGCTGCTGTTGGGCTGGATTAAG-ATACCTTAGACTTGGCAATTTA 14155  
QY 3060 TCAGCAATAGAAATTAATTTGTCACAAATCTGAGGCTGGAATAATCAAGATCAAGCTC 3119  
DB 14156 TAAACATAGAAATTCATTGCTGACAGTTGTGAAGCTGGAAATCAAGATCAAGGCC 14215  
QY 3120 CAGCAGTTCAGTGTCTGCTGAGTGGCTTGTCTGCTTCGAAGATGACCTTTTGGCTGT 3179  
DB 14216 CAGCAATCTGGATCTGCTGATAGGCT--CCCTGCTTCAAAATGGCGCTTCTGCTGC 14273  
QY 3180 GTTCTCA 3186  
DB 14274 ATCTTCA 14280

RESULT 6  
US-09-873-404-3

Sequence 3, Application US/09873404  
Patent No. 6500656  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
FILE REFERENCE: CL001212-CIP  
CURRENT APPLICATION NUMBER: US/09/873,404  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 63588  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(63588)  
OTHER INFORMATION: n = A,T,C or G  
US-09-873-404-3

Query Match 2.5%; Score 80.8; DB 4; Length 63588;  
Best Local Similarity 71.6%; Pred. No. 1.3e-14;  
Matches 106; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 2999 AATGTTTTCAGTCAGTTAGGCTGCTATTAAGAAATATCTTAGAGTGGTATCT 3058  
DB 15658 AATCATCTGTTTATGCTATTCACAGCTATTAACAAATACATTAATGTTGGCTT 15717  
QY 3059 ATCAGCAATAGAAATTAATTTGTTCAAAATCTGAGGCTGGAATAATCAAGATCAAGCT 3118  
DB 15718 ATTAACAGAGAAATTAATTTGTTGAGGCTGGAAGTGGAAAGTCAAGATCAAGGTTG 15777  
QY 3119 CAGCAGTTCAGTGTCTGCTGAGTGGCT 3146  
DB 15778 CAGCAGATTCAGTGTCTGAGGCTT 15805

RESULT 7  
US-09-146-053-4

Sequence 4, Application US/09146053A  
Patent No. 6399349  
GENERAL INFORMATION:  
APPLICANT: Ryan, James W.  
APPLICANT: Sprinkle, Terry Joe Curtis  
APPLICANT: Venema, Richard C.  
TITLE OF INVENTION: Human Aminopeptidase P Gene  
FILE REFERENCE: MCG103  
CURRENT APPLICATION NUMBER: US/09/146,053A  
CURRENT FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/057,854

```
Qy 2983 TCCAGGAACTCCAGAAATGTTTGTTCAGTCAAGTTAGGCTGCT---ATAAGAAAT 3039
Db 68776 TACTAGTAGTGTCAGAACTGATTTGCTTAGTCTGTTTGGCTTTTCATTAACAAAT 68717
Qy 3040 ATTCTTAGTGGGTAACTATATCGCAATAGAAATTTATTTCTCAATCTGAGGCTGG 3099
Db 68716 GCCACAGACTGGGCAATTTACAAACAACAGTAGTTATTTCTCATATGTTCTGAGGCTGG 68657
Qy 3100 AAAATCAGATCAGGCTCCAGCAGGTTCACTGCTGAGGCTGCTGT-TCTGCTTGG 3158
Db 68656 -AGCTCAAGATCAAGCCGCACTAGCTTCACTGCTGAGGCTGCTCTCTGCTTCC 68598
Qy 3159 AAGATGCACTTTTGTCTGTCTCT 3184
Db 68597 AAAATGAGGCTTGTGCTGTCTCT 68572

RESULT 4
US-09-791-211-3/C
; Sequence 3, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Walt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WEN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 3
; LENGTH: 87543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 7421
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7427
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 11609
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 12605
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 12742
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29370
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29422
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29979
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29980
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29981
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30136
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30140
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31205
; OTHER INFORMATION: unknown

NAME/KEY: unsure
LOCATION: 31206
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31592
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33095
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33160
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34066
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34072
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 36816
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
```

OTHER INFORMATION: 99-15668-139.m182, complement  
FEATURE:  
NAME/KEY: primer bind  
LOCATION: 1363..1380  
OTHER INFORMATION: upstream amplification primer  
FEATURE:  
NAME/KEY: primer bind  
LOCATION: 1801..1821  
OTHER INFORMATION: downstream amplification primer, complement  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1489..1513  
OTHER INFORMATION: 99-15668-139 probe  
US-09-539-333D-177

Query Match 2.8%; Score 88; DB 4; Length 3001;  
Best Local Similarity 71.7%; Pred. No. 6.2e-18;  
Matches 129; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

QY 3008 GTTTCAGTCAGTTAGGCTGCTATAGAGAAATCTTAGGCTGCTATCTAGCAAT 3067  
DB 2055 GTTCTAGTCAGTTTGTGCTGCTGTAACAGAAATCCACAGCTGGCTATTAAAGAAAG 2114  
QY 3068 AGGAATTTATTTGTTCAATTTGAGAGCTGGAATAATCCAAGATCCAAGCTCCAGCAGGT 3127  
DB 2115 AAGATTCATTTCTCATAGTTGTGAGAGCTAGAAAGTCCAGATGAAGGCGCCAGCAAT 2174  
QY 3128 TCAGTGTCTGCTGAGTGTCTGTCTGCTTTCGAAGATGGACCTTTTGTGCTGTCTCA 3186  
DB 2175 TTGGTGTGTGTAGGGGTGCGCTGCTCTTCCAAGATGACGCTTTGATGCTGCAATTTCA 2234

## RESULT 2

US-08-781-891-79/c  
Sequence 79, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-Bn  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
TITLE OF INVENTION: WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: NO. 6090620cendurg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
US-08-781-891-79

Query Match 2.7%; Score 86; DB 3; Length 87350;  
Best Local Similarity 73.3%; Pred. No. 3.1e-16;  
Matches 151; Conservative 0; Mismatches 50; Indels 5; Gaps 3;

QY 2983 TCCAGGAACTCCGAATATGTTTGTTCAGTCAAGTCTGCT---ATPAGCAAT 3039  
DB 68776 TACTAGTAGTGCAGAACTGATGTTGCTTAGTCTGTTTCTGTTTCATPACAAAT 68717  
QY 3040 ACTTAGTGGGTATATATATAGCAATAGCAATTTATTTTCAATTTCTGAGGCTG 3099  
DB 68716 GCCACAGACTGGGCAATTTTCAACACAGTAGTTATTTTCATAGTTCTGAGGCTG 68657  
QY 3100 AAAATCAAGATCAAGCTCCAGCAGTTCAGTGTCTGAGTCTGT-TCGCTTCG 3158  
DB 68656 -AGTCCAGATCAAGGCGCCAGTAGTTCAAGTGTGCTGAGGCGCTCTCTCTCC 68598  
QY 3159 AAGATGCACTTTTGTCTGTCTCT 3184  
DB 68597 AAAATGAGGCTTGTGCTGTCTCT 68572

## RESULT 3

US-09-618-166-79/c  
Sequence 79, Application US/09618166  
Patent No. 6583112

GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-Bn  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.

TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
TITLE OF INVENTION: WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/618,166  
FILING DATE: 17-Jul-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052.419c1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-618-166-79

Query Match 2.7%; Score 86; DB 4; Length 87350;  
Best Local Similarity 73.3%; Pred. No. 3.1e-16;  
Matches 151; Conservative 0; Mismatches 50; Indels 5; Gaps 3;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 11:21:38 ; Search time 134.986 Seconds  
(without alignments)  
10417.750 Million cell updates/sec

Title: US-09-900-448-1

Perfect score: 3186  
Sequence: 1 cctgcagctcagcatgct.....accttttgcgtgtcttcca 3186

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfillseq.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	2.8	3001	US-09-539-333D-177	Sequence 177, App
2	86	2.7	87350	US-08-781-891-79	Sequence 79, Appl
3	86	2.7	87350	US-09-618-166-79	Sequence 79, Appl
4	86	2.7	87543	US-09-791-211-3	Sequence 3, Appl
5	82.6	2.6	174493	US-09-804-471A-3	Sequence 3, Appl
6	80.8	2.5	63588	US-09-873-404-3	Sequence 3, Appl
7	80.4	2.5	50000	US-09-146-053-4	Sequence 4, Appl
8	79	2.5	1001	US-09-671-317-212	Sequence 212, App
9	79	2.5	1001	US-08-611-587-1	Sequence 458, App
10	78.4	2.5	8353	US-09-009-913-1	Sequence 1, Appl
11	77.6	2.4	72928	US-09-896-243-344	Sequence 344, App
12	69.8	2.2	762	US-09-322-575-145	Sequence 145, App
13	69.2	2.2	1282	US-09-205-258-197	Sequence 197, App
14	69	2.2	50000	US-09-146-053-3	Sequence 3, Appl
15	68.4	2.1	197496	US-09-877-177A-10	Sequence 10, Appl
16	68.2	2.1	309	US-09-389-681-145	Sequence 145, App
17	68.2	2.1	309	US-09-320-405B-145	Sequence 145, App
18	68.2	2.1	309	US-09-333-826B-145	Sequence 145, App
19	68.2	2.1	309	US-09-333-826B-145	Sequence 145, App
20	68.2	2.1	309	US-09-333-826B-145	Sequence 145, App
21	68.2	2.1	309	US-09-604-287A-145	Sequence 145, App
22	67.8	2.1	50000	US-09-146-053-4	Sequence 4, Appl
23	66.8	2.1	41684	US-09-536-059-1	Sequence 1, Appl
24	66.6	2.1	1001	US-09-641-638-459	Sequence 459, App
25	65.8	2.1	7218	US-08-232-463-14	Sequence 14, Appl
26	63.2	2.0	246240	US-08-724-394A-20	Sequence 20, Appl
27	63.2	2.0	246240	US-08-724-394A-21	Sequence 21, Appl

C 28	63.2	2.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 29	63	2.0	3001	4	US-09-539-333D-192	Sequence 192, App
C 30	62.8	2.0	44453	4	US-09-146-053-5	Sequence 5, Appl
C 31	62	1.9	319608	4	US-09-539-333D-1	Sequence 1, Appl
C 32	62	1.9	319608	4	US-09-679-403-1	Sequence 1, Appl
C 33	61.8	1.9	152331	3	US-09-128-155-16	Sequence 16, Appl
C 34	61.4	1.9	176373	3	US-09-128-155-17	Sequence 17, Appl
C 35	61.4	1.9	8355	3	US-08-406-030A-23	Sequence 23, Appl
C 36	60.8	1.9	168575	4	US-09-426-290-1	Sequence 1, Appl
C 37	60.4	1.9	3001	4	US-09-539-333D-138	Sequence 138, App
C 38	59.4	1.9	202001	4	US-09-734-674-3	Sequence 3, Appl
C 39	59	1.9	80246	3	US-09-078-294-4	Sequence 4, Appl
C 40	59	1.9	80595	3	US-09-078-294-3	Sequence 3, Appl
C 41	57	1.8	49312	4	US-09-671-317-485	Sequence 485, App
C 42	55	1.7	116592	4	US-09-818-512-3	Sequence 3, Appl
C 43	53.6	1.7	49136	3	US-09-422-865-1	Sequence 1, Appl
C 44	52	1.6	43069	4	US-09-292-542A-1	Sequence 1, Appl
C 45	51	1.6	1815	4	US-09-328-475C-339	Sequence 339, App

## ALIGNMENTS

RESULT 1  
US-09-539-333D-177  
Sequence 177, Application US/09539333D  
Patent No. 6476208  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bouguetelert, Lydie  
APPLICANT: Bihain, Bernard  
APPLICANT: Essioux, Laurent  
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS  
FILE REFERENCE: GENSET.047AUS  
CURRENT APPLICATION NUMBER: US/09/539,333D  
CURRENT FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/126,903  
PRIOR FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: US 60/131,971  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/132,065  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/143,928  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: US 60/145,915  
PRIOR FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: US 60/146,453  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/146,452  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/162,288  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: US 09/416,384  
PRIOR FILING DATE: 1999-10-12  
NUMBER OF SEQ ID NOS: 231  
SOFTWARE: Patent.pm  
SEQ ID NO 177  
LENGTH: 3001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 1501  
OTHER INFORMATION: 99-15668-139 : polymorphic base C or T  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1482..1500  
OTHER INFORMATION: 99-15668-139.mis1  
FEATURE:  
NAME/KEY: misc binding  
LOCATION: 1502..1521

Db 62 AGAGCTTCCTGAGCTTGGGACGAGATTGAAGTGGAGATGTGGCCAAAGCCGATCCAGAA 121  
 Qy 150 GTGACTGAACGCTGCTCAGATGGCTGGAGCTTTGATGCTCAACCCCTGATGAACAATGGA 209  
 Db 122 GTACCGAAGCGCTGCTGAGATGGCTGGAGCTTTGATGCTCAACCCCTGATGAACAATGGA 181  
 Qy 210 ACCATGCTGTTTTTAAAGGGGAGTTTGTGTGGAAGATGCAACAATGGGACCGGAGTTA 269  
 Db 182 AACATGCTGTTTTTAAAGGGGAGTTTGTGTGGAAGATGCTGCTGGGCGCGGAGTTG 241  
 Qy 270 ATCTCAGAGAGATGGAAGATTTTCCCAAGCCCTGTGATGCTGATTCGCTCAAGTGCAC 329  
 Db 242 ATCTCAGAGAGATGGAAGATTTTCCCAAGCCCTGTGATGCTGATTCGCTCAAGTGCAC 301  
 Qy 330 AACAGTGTCTTTCTGATCAAGGGGAGCAAAAGTGTGGTATACCTCCTGAAAAGAGAG 389  
 Db 302 AACAGTGTCTTTCTGATCAAGGGGAGCAAAAGTGTGGTATACCTCCTGAAAAGAGAG 358  
 Qy 390 AAAAGATACCCAAAGTGTCTCAAGATGAATTTCTGGAATCCCATCCCACTGATGCA 449  
 Db 359 GAGAGATATCCAAAGTGTCTCAAGATGAATTTCTGGAATCCCATCCCACTGATGCA 417  
 Qy 450 GCTGTGGA 457  
 Db 418 GCTGTGGA 425

## RESULT 15

ABX44213  
ID ABX44213 standard, cDNA; 425 BP.

AC ABX44213;

DT 21-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #9378.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;

XX muscle deposition; fat deposition; genome mapping; gene identification;

XX Gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-0960352.

XX 12-JAN-1999; 99US-115707P.

XX 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI, 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat

XX deposition, useful for genome mapping, gene identification and

XX analysis, cattle breeding, or for genetically improving cattle

XX Claim 2; SEQ ID No 9378; 245BP; English.

CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 1512 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridization between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression or for genetically  
 CC improving cattle. The present sequence is one of the 1512 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the USPTO web site:  
 CC <http://www.uspto.gov/sequence.html?docid=20020137139>.  
 CC Seqdata.uspto.gov/sequence.html?docid=20020137139.

Query Match 8.1%; Score 259; DB 25; Length 425;  
 Best Local Similarity 79.1%; Pred. No. 2, 4e-72;  
 Matches 336; Conservative 0; Mismatches 80; Indels 9; Gaps 2;

Qy 6 CAGCTCAGCATGCTGCTGAGGATCTGAGGACACCCCTGCTCACTGGGGTTGTGAGACCTATGC 65  
 Db 4 CAGCTCAGCATGCTGCTGAGGATCTGAGGACACCCCTGCTCACTGGGGTTGTGAGACCTATGC 63  
 Qy 66 TGGTCTCTGGCCATGTCACACCCCTCTTCTCCGACTAGTGCAGATGGAGAAATGTTGCTGAA 125  
 Db 64 TGGTCTCTGGCCATGTCACACCCCTCTTCTCCGACTAGTGCAGATGGAGAAATGTTGCTGAA 123  
 Qy 126 GGGGAGA-----CAAGCCAGACCCAGACGTGATGAAAGCTGCTCAATGCTGAGC 179  
 Db 124 GGTGGAAATGTGGCCAGACCCAGACGTGATGAAAGCTGCTCAATGCTGAGC 183  
 Qy 180 TTTGATGCTACACCCCTGATGACAAATGGAACCATGCTGTTTTTAAAGGGAGTTGTG 229  
 Db 184 TTTGATGCTACACCCCTGATGACAAATGGAACCATGCTGTTTTTAAAGGGAGTTGTG 243  
 Qy 240 TGGAAAGTCAAAATGGGAGCGGAGTTAATCTCAGAGATGAAAGAAATTTCCCAAGC 299  
 Db 244 TGGAAAGTCAAAATGGGAGCGGAGTTAATCTCAGAGATGAAAGAAATTTCCCAAGC 303  
 Qy 300 CCTGTGATGCTGATTCCTGCTCAAGGTCAACAGTGTCTTTCTGATCAAGGGGAGCAAA 359  
 Db 304 CCTGTGATGCTGATTCCTGCTCAAGGTCAACAGTGTCTTTCTGATCAAGGGGAGCAAA 363  
 Qy 360 GTTGTGGTATACCTCTCTGAAAAGAGAAAGATGCCAAAGTTGCTCCAAATGAA 419  
 Db 364 TTTCTGGGTATATCTCTCTG---AGAAAGGGGAGAAATATCCAAAGTTGCTCCAAAGAG 420  
 Qy 420 TTTCC 424  
 Db 421 TTTCC 425

Search completed: December 15, 2003, 12:36:47  
 Job time : 508.03 secs



XX Claim 2; SEQ ID No 445; 245bp; English.

PS The invention relates to a purified nucleic acid molecule associated with  
XX laccation or muscle and fat deposition (designated LMPD), derived  
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMPD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.

XX Sequence 405 BP; 93 A; 91 C; 130 G; 91 T; 0 other;

Query Match 8.2%; Score 261.8; DB 25; Length 405;  
Best Local Similarity 81.3%; Pred. No. 2.9e-73;  
Matches 331; Conservative 0; Mismatches 67; Indels 9; Gaps 2;

75 GCCATTGCCACCCCTCTCCGACGAGCCCATGAGGAAATTTGCTAAGGCGAGA-- 132  
2 GCTTAAGCCCACTCTGCTGCTAGAGCTCTGAGCTTGGGACGAGTTGAAGTGGAAAT 61  
133 -----CCAAGCCAGACCCAGAGCTGAGTGAACGCTGCTCAGATGCTGAGAGCTTGAATGCT 188  
62 GTGGCCAGGCGGATCCAGAAAGTGAAGGAAAGCTGCTGGAGCTTGAATGCT 121  
189 ACACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 248  
122 ACCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181  
249 CACAAATGGGAGCGGAGTGAATCTCAGAGATGAGAAATTTCCCGACCTGAGAT 308  
182 CATGCTGGGCGCGGAGTGAATCTCAGAGATGAGAAATTTCCCGACCTGAGAT 241  
309 GCTGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 368  
242 GCTGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301  
369 TACCTCTCTGAAAAGAGAGAAAGATACCAAGTTGCTCCAGATGAATTTCTGGA 428  
302 TATCTCTCTG---AGAAAGGAGGAGAAATTCCAAGTTGCTCCAGATGAATTTCTGGA 358  
429 ATCCCATCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 475  
359 ATCCCATCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405

RESULT 14

ID ABX38537 standard; cDNA; 425 BP.  
XX ABX38537,  
XX 20-FEB-2003 (first entry)

XX Bovine EST associated with laccation/muscle/fat deposition #3702.

DB Bovine, 89; EST; expressed sequence tag; laccation; LMPD;  
XX muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001, 2001US-0960352.

XX 12-JUN-1999, 99US-115707P.

XX 11-JAN-2000, 2000US-0480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAN/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-110599/10.

XX Claim 2; SEQ ID No 3702; 245bp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
XX laccation or muscle and fat deposition (designated LMPD), derived  
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 15112 nucleotide  
CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMPD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 15112 bovine  
CC LMPD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?docid=20020137139.

XX Sequence 425 BP; 88 A; 95 C; 138 G; 103 T; 1 other;

Query Match 8.2%; Score 260.2; DB 25; Length 425;  
Best Local Similarity 80.4%; Pred. No. 9.7e-73;  
Matches 344; Conservative 0; Mismatches 74; Indels 10; Gaps 3;

36 CCCGTTGCACTGGGCTTGGAGCCCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 95  
2 CCAAGTGAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61  
96 CCAAGTGAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149

Matches 308; Conservative 0; Mismatches 53; Indels 3; Gaps 1;

QY 145 CAGACGTGACTGAACGCTGCTCAGATGCTGAGCTTGTGATGCTACCACTGGATGACA 204

DB 1 CAGAGTGAAGGAAACGCTGCTCGGATGCTGGGCTTTGATGCTACCACTGGATGAGC 60

QY 205 ATGGAACATGCTGCTTTTAAAGGGGAGTTGTGTGGAAGATCACAAATGGAGCCGGG 264

DB 61 ATGGGAACATGCTGCTTTTAAAGGGGAGTTGTGTGGAAGATCACAAATGGAGCCGGG 120

QY 265 AGTTAACTCAGAGAGATGGAAGAAATTTCCCAAGCCCTGAGATGCTGCAATTCGCTGAG 324

DB 121 AGTTGATCTCAGAGAGATGGAAGATTCCTCCAGTCCCGTGAATGCTGCAATTCGCTGAG 180

QY 325 GTCAACAACATGCTTTTCTGATCAAGGGGAGCAAAAGTCTGAGTACCTCTGGAAGAA 384

DB 181 ATGTAAACATGCTGCTTTCTGATCAAGGGGAGCAAAATTCCTGAGTATACCTCTGAG 237

QY 385 AGGAGAAAGATACCCCAAGTTGCTCCAAATGATGATTTCTGGAATCCCATCCCACTGG 444

DB 238 AGGGGGAGGAATATCCAAAGTTGCTCCAAAGATGATTTCTGGAATCCCATCCCACTGG 297

QY 445 ATGCAAGCTGGAATGTCACCGTGGAGATGTCAAGCTGAAGGCTCTCTCTTCTGCAAG 504

DB 298 ATGCAAGCTGGAATGTCACCGTGGAGATGTCTCATGAGGATGTCCTCTCTTCTGCAAG 357

QY 505 GCCA 508

DB 358 GCAA 361

RESULT 12  
AAC00310  
ID AAC00310 standard; cDNA; 279 BP.

XX AAC00310;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 308.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX BP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GSEST) GENSET.

XX Dumas Maline Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG00304.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 308; 71bp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. An ORF has been identified within the  
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
XX derived from 30 different tissues. EST sequences usually correspond  
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are  
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

XX Sequence 279 BP; 56 A; 71 C; 87 G; 64 T; 1 other;

Query Match 8.3%; Score 263.6; DB 21; Length 279;  
Best Local Similarity 99.6%; Pred. No. 5; 9e-74;  
Matches 263; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTGAGCTCAGCATGAGGCTAGGCTAGGAGACACCCGTTGCACTGGGTTGTGAGCC 60

DB 16 CTGTGAGCTCAGCATGAGGCTAGGCTAGGAGACACCCGTTGCACTGGGTTGTGAGCC 75

QY 61 TATGCTGCTCTGAGCATTTGCCACCTCTTCTCCGACTAGTGCCTGAGGAATGTTG 120

DB 76 TATGCTGCTCTGAGCATTTGCCACCTCTTCTCCGACTAGTGCCTGAGGAATGTTG 135

QY 121 CTGAAGCCGAGACCAAGCCGAGACCTGAGTGAAGCTGCTGATGCTGGAAGCT 180

DB 136 CTGAAGCCGAGACCAAGCCGAGACCTGAGTGAAGCTGCTGATGCTGGAAGCT 195

QY 181 TTGATGCTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

DB 196 TTGATGCTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255

QY 241 GGAAGATCACAATGGAAGCCGGG 264

DB 256 GGAAGATCACAATGGAAGCCGGG 279

RESULT 13  
ABX35280  
ID ABX35280 standard; cDNA; 405 BP.

XX ABX35280;

XX 20-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #445.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMD;

XX muscle deposition; fat deposition; genome mapping; gene identification;  
XX gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-0960352.

XX 12-JAN-1999; 99US-115707P.

XX 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat  
XX deposition, useful for genome mapping, gene identification and  
XX analysis, cattle breeding, or for genetically improving cattle

XX PI Byatt JC, Mathalagan N, Tao N, Warren WC;  
 XX MPI, 2003-110599/10.  
 DR New nucleic acid associated with lactation, and muscle and fat  
 XX deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 XX  
 PS Claim 2; SEQ ID No 6846; 245bp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
 CC  
 SQ Sequence 383 BP; 85 A; 108 C; 94 G; 96 T; 0 other;  
 Query Match 8.4%; Score 266.8; DB 25; Length 383;  
 Best Local Similarity 82.1%; Pred. No. 6,7e-75;  
 Matches 307; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 QY 555 CCTGACTATATGCGCTGATGAGCCCACTATGCTTGTCTGCACTGACGCTGCAACCAT 614  
 DB 10 CATATGATATGAGCGCTGATGAGCCCACTATGCTTGTCTGCACTGACGCTGCAACCAT 69  
 QY 615 GGTGCAACCTATGCGCTGATGAGCCCACTATGCGCTTGTCTGCACTGACGCTGCAACCAT 674  
 DB 70 AGTGCACCTATGCGCTTGTAGTGAACCACTATGCGCTTGTCTGCACTGACGCTGCAACCAT 129  
 QY 675 TGGCATAGCTGAGCCATGCTCTCATGATGAGCCCACTATGCGCTTGTCTGCACTGACGCTGCAACCAT 734  
 DB 130 TGGCACAAGCTGCTATGATGATCTGTGCGCCCAAGTCCCTCAACAGTGAATGCTGCC 189  
 QY 735 TTTTCTGGAGAGAAAAGTCTATGCTGTCAGGAGCAACCAAGTATATGCTTCTGACA 794  
 DB 190 TTTTCTGGAGATTAACGCTCTTCACTATCAGGAGTACCAAGTATATATCTTCTGACA 249  
 QY 795 AAGGGAGGCTATACCTAGTATAGCGGTTATCCGAACCGCTGAGAGAAAGTCCGGAGAC 854  
 DB 250 AAGGAGGCTATACCTAGTATAGCGGTTATCCGAACCGCTGAGAGAAAGTCCGGAGAC 309  
 QY 855 CCTCATGAGGATTAATCCTGAGACTCTGAGATGCGGCGCTTATCTGCGGCTTCTCGG 914  
 DB 310 CCGAGAGGGGCTGCGCTTCTGATCTGTGAGTGAAGCGCTTAACTGCTGATCTTCTCAG 369  
 QY 915 CTCCTATATCATGGC 928  
 DB 370 CTCCTATATCATGGC 383

RESULT 11  
 ABX35215  
 ID ABX35215 standard; cDNA; 437 BP.  
 XX  
 AC ABX35215;  
 XX  
 XX 20-FEB-2003 (first entry)  
 DT  
 XX  
 DE Bovine EST associated with lactation/muscle/fat deposition #280.  
 XX  
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 KM muscle deposition; fat deposition; genome mapping; gene identification;  
 KM gene analysis; cattle breeding.  
 XX  
 OS Bos Taurus.  
 OS  
 PN US2002137139-A1.  
 PN  
 PD 26-SEP-2002.  
 XX  
 XX 24-SEP-2001; 2001US-0960352.  
 PF  
 XX 12-JAN-1999; 99US-115707P.  
 PR  
 PR 11-JAN-2000; 2000US-0480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 PI Byatt JC, Mathalagan N, Tao N, Warren WC;  
 XX MPI, 2003-110599/10.  
 DR New nucleic acid associated with lactation, and muscle and fat  
 XX deposition, useful for genome mapping, gene identification and  
 PT analysis, cattle breeding, or for genetically improving cattle  
 PT  
 XX  
 PS Claim 2; SEQ ID No 380; 245bp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMPD), derived  
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a  
 CC second nucleic acid molecule comprising any of 15112 nucleotide  
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.  
 CC Also included are; (1) a transformed cell having a nucleic acid  
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 CC translated sequence that functions in the cell to cause termination of  
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
 CC of the mRNA molecule; and (2) determining a level or pattern of a  
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its  
 CC complement or fragment) with a complementary nucleic acid molecule  
 CC obtained from the bovine cell or tissue, where hybridisation between the  
 CC marker nucleic acid and the complementary nucleic acid permits the  
 CC detection of the molecule; and (b) detecting the level or pattern of the  
 CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMPD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 15112 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?docid=20020137139.  
 CC  
 SQ Sequence 437 BP; 100 A; 102 C; 128 G; 104 T; 3 other;  
 Query Match 8.3%; Score 264.4; DB 25; Length 437;  
 Best Local Similarity 84.6%; Pred. No. 4.4e-74;

QY 1007 GGTAGACGAGCCCTTGTATGAGAAAGTCCCTGAGCCCTTACATGATCCGCCAATGG 1066  
 DB 241 AGTCGATGGGGCCCTGTGTACAGAGAGTCTCTGGGCCCACTCGGTTCTGCCAATGG 300  
 QY 1067 TCCCGGCTTACTCATCTCATGAGTGTCCCAATTTGTACTGCTACAGATGTGAGAAACT 1126  
 DB 301 TCTGGGCTGTACTCTGTCCAGGCCCAATCTGTACTGCTACAGAAATGTGAGAAAT 360  
 QY 1127 GAATGACGACGAGCCCTTCCGCAAGCCCAAGAAATGTGACAGTCTCT 1174  
 DB 361 GAGCAAGACCAAGACCTTCCCGAGGCCCAAGAGATGAACAGCTCT 408

RESULT 9  
 ID ABX40203 standard; cDNA; 374 BP.  
 AC ABX40203;  
 DE 20-FEB-2003 (first entry)  
 XX Bovine EST associated with lactation/muscle/fat deposition #5368.  
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 XX muscle deposition; fat deposition; genome mapping; gene identification;  
 XX gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 XX US2002137139-A1.  
 XX 26-SEP-2002.  
 XX 24-SEP-2001; 2001US-0960352.  
 XX 12-JAN-1999; 99US-115707P.  
 XX 11-JAN-2000; 2000US-0480902.  
 XX  
 XX (BYAT/) BYATT J C.  
 XX (MATH/) MATHIALAGAN N.  
 XX (TAON/) TAO N.  
 XX (WARR/) WARREN W C.  
 XX  
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX MPI; 2003-110599/10.  
 XX  
 XX New nucleic acid associated with lactation, and muscle and fat  
 XX deposition, useful for genome mapping, gene identification and  
 XX analysis, cattle breeding, or for genetically improving cattle  
 XX  
 XX Claim 2; SEQ ID No 5368; 245bp; English.  
 XX  
 XX The invention relates to a purified nucleic acid molecule associated with  
 XX lactation or muscle and fat deposition (designated LMPD), derived  
 XX from cattle, and the LMPD nucleic acid can specifically hybridize to a  
 XX second nucleic acid molecule comprising any of 1512 nucleotide  
 XX sequences, appearing as ABX34836-ABX49947, or complements of them.  
 XX Also included are; (1) a transformed cell having a nucleic acid  
 XX comprising an LMPD nucleic acid linked to a promoter and a 3' non-  
 XX translated sequence that functions in the cell to cause termination of  
 XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
 XX of the mRNA molecule; and (2) determining a level or pattern of a  
 XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 XX nucleic acid (comprising any of the 1512 nucleic acid sequences or its  
 XX complement or fragment) with a complementary nucleic acid molecule  
 XX obtained from the bovine cell or tissue, where hybridization between the  
 XX marker nucleic acid and the complementary nucleic acid permits the  
 XX detection of the molecule; and (b) detecting the level or pattern of the  
 XX complementary nucleic acid, where the detection of the complementary  
 XX nucleic acid is predictive of the level or pattern of the molecule.  
 XX The LMPD nucleic acid is used for determining a level or pattern

CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 1512 bovine  
 CC LMPD EST (expressed sequence tag) nucleic acids.  
 CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC [seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139).  
 XX  
 SQ Sequence 374 BP; 86 A; 102 C; 92 G; 94 T; 0 other;  
 Query Match 8.4%; Score 269.2; DB 25; Length 374;  
 Best Local Similarity 83.0%; Pred. No. 1.1e-75;  
 Matches 307; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 555 CCTGATATATAGGCTGTAGAGCCCACTATAGTCTTGTCTGCACTGACGTGTGCAACCAT 614  
 DB 5 CATATGATTAAGGCTGTAGAGCCCACTATAGTCTTGTCTGCACTGCTTTCTGCAACCAT 64  
 QY 615 GGTGCCACCTATAGCCTTTCAGTGGAGACCACTACTGAGCTGTGACACAGCCGGATGGC 674  
 DB 65 AGTGCACCTTACGCTTTTCAGTGGAGACCACTACTGAGCTGTGACACAGCCGGAGCGG 124  
 QY 675 TGGCATAGCTGGGCCATTGCTCATCATGAGTGGGCCCAAGGTCCTTTCAGCAGTGTGCTGC 734  
 DB 125 TGGCAGAGCTGGGGCATGTAGCATCTGTGGCCCAAGTCCCTCAACAGTGTGCTGCC 184  
 QY 735 TTTTCTGGGAGAAAGAAATCTATCTGTGTCAGGAGCAACCAAGTATATGCTTCTCGACA 794  
 DB 185 TTTTCTGGGAGTAAAGGCTATCTATCTGATCAAGGAGTACCAAGTATATATCTTCTCGACA 244  
 QY 795 AAGGAGGCTATATACCTAGTAAGCGGTATCCGAAGCGGCTGGAGAGAGTGGGAGCC 854  
 DB 245 AGGCGAGGCTTACACTCTCTGTAAGATTTATCCAAAGCACTGGAGAGAGATTTGGAGC 304  
 QY 855 CCTGATGAGATTAATCTGAGCTGTGTGATGCGGCTTATCTGCTCGGCTTCTCGG 914  
 DB 305 CCGATGGGGGTCTGAGCTTATCTGTGATGAGTACGCTTACGCTGATCTTCTCGAG 364  
 QY 915 CTTCATATCA 924  
 DB 365 CTCTACATCA 374

RESULT 10  
 ID ABX41681 standard; cDNA; 383 BP.  
 AC ABX41681;  
 DE 20-FEB-2003 (first entry)  
 XX Bovine EST associated with lactation/muscle/fat deposition #6846.  
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;  
 XX muscle deposition; fat deposition; genome mapping; gene identification;  
 XX gene analysis; cattle breeding.  
 OS Bos Taurus.  
 XX  
 XX US2002137139-A1.  
 XX 26-SEP-2002.  
 XX 24-SEP-2001; 2001US-0960352.  
 XX 12-JAN-1999; 99US-115707P.  
 XX 11-JAN-2000; 2000US-0480902.  
 XX  
 XX (BYAT/) BYATT J C.  
 XX (MATH/) MATHIALAGAN N.  
 XX (TAON/) TAO N.  
 XX (WARR/) WARREN W C.

comprising an LMPD nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 15112 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridization between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMPD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 15112 bovine LMPD EST (expressed sequence tag) nucleic acids.

Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site: [seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139).

Sequence 420 BP; 90 A; 114 C; 116 G; 100 T; 0 other;

Query Match 9.4%; Score 298.4; DB 25; Length 420;  
Best Local Similarity 81.9%; Pred. No. 4.6e-85;  
Matches 344; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

663 AGCCGGGATGCTGGCATAGCTGGCCCATTCATCATAGTGGCCCAAGGCTCTTAGCA 722  
1 AGCCGGGAGCGGGTGGCAAGCTGCTCATTCATTCATTCGCGCCCAAGGCTCTCAACA 60  
723 GGGATGCTGCTCTTCTTCTGGGAGAGAAAATCTATCTGCTCCAGGGCAACCGGATAT 782  
61 GGGATGCTGCTCTTCTTCTGGGATTAAGAGCTTACCTCATCCAGGATACCGGATAT 120  
783 GCTTCTCTGATCAAGAGGAGCTATACCTATAGAGGATTAAGAGGCTGAGAG 842  
121 ATCTTCTGATCAAGAGGAGCTATACCTATAGAGGATTAAGAGGCTGAGAG 180  
843 GAAGTGGGAGCCCTCATGAGGATTAATCTGAGACTTGTGAGAGCGCTTATCTGCCCT 902  
181 GAATTTGGAGGCGCTGATGGGCTGCTCTTCTGATGATGACAGCTTATCTGCTCT 240  
903 GGGTCTTCTGCTCATATCATGAGAGAGCGCGCTGAGGCTGAGACTGAGAGTCA 962  
241 GATCTTCTGAGCTCATATCATGAGAGAGCGCAAGAGCTGAGAGCTGAGAGTCA 300  
963 GAGCCCAAGCCAGCTGAGAGAGCTTCTTGGCCCATGAGAGGATGAGAGGCTTG 1022  
301 GAGGCTCAAGCCAGCTGAGAGAGCTTCTTGGCTCATGAGAGGATGAGAGGCTTG 360  
1023 TGTATGAGAAAGTCCCTTGGCCCTTAATCATATGTTCCGCAATGATGCCGCTTGAACCTT 1082  
361 TGTATGAGAAAGTCTTGGCCCTTCAATGATGATGATGATGATGATGATGATGATGAT 420

RESULT 8

ABX41875  
ID ABX41875 standard; cDNA; 408 BP.

XX  
XX  
AC ABX41875;

XX  
XX  
DT 20-FEB-2003 (first entry)

XX  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #7040.

XX  
XX  
KM Bovine; ss; EST: expressed sequence tag; lactation; LMPD;

XX  
XX  
KM muscle deposition; fat deposition; genome mapping; gene identification;

XX  
XX  
KM gene analysis; cattle breeding.

XX  
OS  
XX Bos Taurus.

PN US2002137139-A1.  
XX  
XX 26-SEP-2002.  
PD  
XX  
XX 24-SEP-2001; 2001US-0960352.  
PF  
XX  
XX 12-JAN-1999; 99US-115707P.  
PR 11-JAN-2000; 2000US-0480902.  
XX  
XX (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
PI  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI, 2003-110599/10.  
XX  
XX  
XX New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and  
PT analysis, cattle breeding, or for genetically improving cattle -  
XX  
XX Claim 2; SEQ ID NO 7040; 245bp; English.

The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated LMPD), derived from cattle, and the LMPD nucleic acid can specifically hybridise to a second nucleic acid molecule comprising any of 15112 nucleotide sequences, appearing as ABX34836-ABX49947, or complements of them. Also included are: (1) a transformed cell having a nucleic acid comprising an LMPD nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 15112 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridization between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMPD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 15112 bovine LMPD EST (expressed sequence tag) nucleic acids.

Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site: [seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139).

Sequence 408 BP; 95 A; 109 C; 111 G; 93 T; 0 other;

Query Match 8.8%; Score 280; DB 25; Length 408;  
Best Local Similarity 80.4%; Pred. No. 3.9e-79;  
Matches 328; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

767 GGGACCCAGGATATGCTTCTTCTGACAAAGGAGGCTATACCTAGTAAGCGGTTATCC 826  
1 GGGTACCCAGGATATATCTTCTGACAAAGGAGGCTATACCTAGTAAGGATATCC 60  
827 GAAGCGGCTGAGAGAGAGTGGGACCCCTCATATGAGGATTAATCTGAGACTCTGAGATCC 886  
61 AAAGCAGCTGAGAGAGAGATTTGGAGCCCTATATGGGCTCTCTTCATTTCTGTGATGC 120  
887 GGCCTTATCTGCTCCCTGAGGCTTCTGAGCTCATATCATAGGAGAGCGGCGCTGTGATG 946  
121 AGCTTTACTGCTCTGAGATCTTCTGAGCTCATATCATAGGAGAGCGGAGAGCTGTGAG 180  
947 GCTGACCTGAGAGTCAAGAGCCCAAGCCAGTGAACAGAGCTTCTTGGCCCATGAGAA 1006  
181 GCTGACCTGAGAGTCAAGAGCTCAAGCCAGTGAACAGAGCTTCTTGGCTCATATGAA 240

Db 562 AGCTTG 567

## RESULT 6

ABN96669/c  
ID ABN96669 standard; DNA; 504 BP.

ABN96669;

13-AUG-2002 (first entry)

Gene #3167 used to diagnose liver cancer.

Gene: liver cancer; ds: hepatocellular carcinoma; hepatotropic;  
KM metastatic liver tumour; cytostatic; expression profile; disease state;  
KM disease progression; drug toxicity; drug efficacy; drug metabolism.

Homo sapiens.

MO200229103-A2.

11-APR-2002.

02-OCT-2001; 2001MO-US30589.

02-OCT-2000; 2000US-237054P.

(GENE-) GENE LOGIC INC.

Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

WPI; 2002-426119/45.

Diagnosing and detecting the progression of liver cancer,  
hepatocellular carcinoma or metastatic liver tumor in a patient,  
involves detecting the level of expression of two or more genes in a  
liver tissue sample -

Claim 1; SEQ ID NO 3167; 298bp; English.

The invention relates to a novel method for diagnosing and detecting the  
progression of liver cancer, hepatocellular carcinoma or metastatic liver  
tumour in a patient, and differentiating metastatic liver cancer from  
hepatocellular carcinoma in a patient, involving detecting the level of  
expression of two or more genes represented in ABN93503-ABN97455 in a  
tissue sample. The method of the invention has hepatotropic, and  
cytostatic activity. The method is useful for diagnosing and detecting  
the progression of liver cancer, hepatocellular carcinoma and metastatic  
liver carcinoma in a patient. The method is useful for identifying  
expression profiles which serve as useful diagnostic markers as well as  
markers that can be used to monitor disease states, disease progression,  
drug toxicity, drug efficacy and drug metabolism.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 504 BP; 123 A; 119 C; 138 G; 110 T; 14 other;

Query Match 9.8%; Score 313.4; DB 24; Length 504;  
Best Local Similarity 91.4%; Pred. No. 7.5e-50;  
Matches 468; Conservative 0; Mismatches 29; Indels 15; Gaps 13;

833 GAAGAGAGTCGAGACCCCTCATGATGATCTCTGAGCTCTGTGAGACCGCTTTATCTG 898

498 GAAGAAGAGTGAACCCCTCATGATGATCTCTGAGCTCTGTGAGACCGCTTTATCTG 440

899 CCTGAGTCTTCTGAGCTCATGATGATGAGAGCGGCTGTGAGCTGTGAGACCGTGA 958

439 CCTT-GCTTCTTCTGAGCTCATGATGATGAGAGCGGCTGTGAGCTGTGAGACCGTGA 381

959 GTGAGAGCCCAAGCAGCTGTGAGAGCTTCTTGG-CCCATGAGAGAGTGAACGAG 1017

Db 380 GTGAGAG-NCCAAGCAGCTGTGAGAGAGCTTCTTGGACCCCATGAGAAAGTAGAGAG 322

QY 1018 CCTGTGTATGAGAAAGTCCCTTGGCCCTTAACATGATTCGCGCATGTCGGGCTGT 1077

Db 321 CTTGTGTATGAGAAAGTCCCTTGGCCCTTAACATGATTCGCGCATGTCGGGCTGT 263

QY 1078 ACCTCATTCATGAGTCCCAATTTGTACTGTACAGTATGTGAGAAAGTGAATGACCA 1137

Db 262 ACCTCATTCATGAGTCCCAATTTGTACTGTACAGTATGTGAGAAAGTGAATGACCA 205

QY 1138 AGGCCCCCTTCGCAACCCCAAGATGTGACCAAGTCTCTGGGCTGCACTGAGGGGCT 1197

Db 204 AGGCCCCCTTCGCAACCCCAAGATGTGACCAAGTCTCTGGGCTGCACTGAGGGGCT 146

QY 1198 TCTGATATGATGAGTCCCTGCGCCCACTTCATGATTCCTCATATTAAGACATGCTT 1257

Db 145 TCTGATATGATGAGTCCCTGCGCCCACTTCATGATTCCTCATATTAAGACATGCTT 89

QY 1258 CTTGCTTCTCACTGAGGGGCTTCTGATGATGATGAGTCTGAGCCCACTCCCAAGTT 1317

Db 88 CTT-GCTTCTCACTGAGGGGCTTCTGATGATGATGAGTCTGAGCCCACTCCCAAGTT 33

QY 1318 TCTCATATTAAGACAGATTTCTTCTCACTT 1349

Db 32 TCTCATATTAAGACAGATTTCTTCTCACTT 1

## RESULT 7

ABX45158  
ID ABX45158 standard; cDNA; 420 BP.

ABX45158;

21-FEB-2003 (first entry)

Bovine EST associated with lactation/muscle/fat deposition #10323.

Bovine; ss; EST; expressed sequence tag; lactation; LMPD;

KM muscle deposition; fat deposition; genome mapping; gene identification;

KM gene analysis; cattle breeding.

Bos Taurus.

US2002137139-A1.

26-SEP-2002.

24-SEP-2001; 2001US-0960352.

12-JAN-1999; 99US-115707P.

11-JAN-2000; 2000US-0480502.

(BYAT/) BYATT J C.

(MATH/) MATHIALAGAN N.

(TAON/) TAO N.

(WARR/) WARREN W C.

Byatt JC, Mathialagan N, Tao N, Warren WC;

WPI; 2003-110599/10.

New nucleic acid associated with lactation, and muscle and fat

deposition, useful for genome mapping, gene identification and

analysis, cattle breeding, or for genetically improving cattle -

Claim 2; SEQ ID NO 10323; 245bp; English.

The invention relates to a purified nucleic acid molecule associated with

lactation or muscle and fat deposition (designated LMPD), derived

from cattle, and the LMPD nucleic acid can specifically hybridize to a

second nucleic acid molecule comprising any of 1512 nucleotide

sequences, appearing as ABX4836-ABX49947, or complements of them.

Also included are; (1) a transformed cell having a nucleic acid





DB 1482 TTCCTCATATAAAGACAGATTGCTTCTGCTTCACAGAGGGCCCTTCGATGAG 1541  
 QY 1292 TCTGGCTTGGCCCGCCAGCTCCCGATTTCATATAAAGACAGATTGCTTCTGATGA 1351  
 DB 1542 TCTGGCTTGGCCCGCCAGCTCCCGATTTCATATAAAGACAGATTGCTTCTGATGA 1601  
 QY 1352 ATCAA 1356  
 DB 1602 AAAAA 1606

RESULT 3  
 AB283572  
 ID AB283572 standard; cDNA; 690 BP.  
 AC AB283572;  
 XX 14-MAY-2003 (first entry)  
 DT Toxicologically relevant human nucleotide sequence #731.  
 XX Toxicologically relevant gene; toxicological response; gene; ss.  
 KM Homo sapiens.  
 OS MO2003016500-A2.  
 XX 27-FEB-2003.  
 PD 16-AUG-2002; 2002WO-US26514.  
 XX 16-AUG-2001; 2001US-313080P.  
 PR (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.  
 PA Neft RE, Dunn RT, Adkins K, Pickett GG, Kier LD, Schmeisler K;  
 PI Alen P;  
 XX WPI; 2003-268322/26.  
 DR Determining a toxicological response to an agent, useful for screening  
 PT of drugs, comprises comparing the expression profile of one or more  
 PT human toxic response genes to a reference gene expression profile  
 PT indicative of toxicity -  
 PS Claim 1; Page 241; 455pp; English.  
 XX The present invention describes a method (M1) for determining a  
 CC toxicological response to an agent, which comprises comparing the  
 CC expression profile of one or more human toxic response genes to a  
 CC reference gene expression profile indicative of toxicity, and so  
 CC determining the presence of a toxic response to the agent. Also  
 CC described: (1) an array comprising one or more polynucleotides selected  
 CC from the genes corresponding to the partial sequences given in AB282842  
 CC to AB284764, or their fragments of at least 20 nucleotides, or  
 CC homologues; and (2) determining if a gene putatively identified to be a  
 CC toxic response gene plays a role on toxic response pathways by  
 CC determining the expression profile of the gene after exposure of cells  
 CC or a human subject to a known toxic pharmaceutical or industrial agent,  
 CC comprising: (a) exposing cells to an agent or isolating cells from a  
 CC human subject who was exposed to an agent; (b) obtaining the test gene  
 CC expression profile for a putatively identified toxic response gene after  
 CC exposure to a known toxic pharmaceutical or industrial agent; and  
 CC (c) comparing the test profile to the expression profile of a gene with  
 CC a similar function or comparing the test profile to the expression  
 CC profile of that gene after exposure to other known toxic compounds. The  
 CC methods are useful for predicting and determining toxicological responses  
 CC on a cellular, organ or system level. The arrays comprising the human  
 CC genes are useful for toxicological screening of drugs, pharmaceutical  
 CC compounds and chemicals.  
 XX Sequence 690 BP; 147 A; 181 C; 196 G; 164 T; 2 other;

Query Match 16.0%; Score 509.4; DB 25; Length 690;  
 Best Local Similarity 99.8%; Pred. No. 5.5e-153;  
 Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 630 TTCAGTGGAGACCCACTACTAGCGGTCTGACACACAGCCGGGATGGCTGAGTACGGCC 689  
 DB 72 TTCAGTGGAGACCCACTACTAGCGGTCTGAGCACACAGCCGGGATGGCTGAGTACGGCC 131  
 QY 690 ATTGCTCATCAGTGGCCCGAGGGTCTCTGACAGTGGATGTCCTTTCTTGGAGAA 749  
 DB 132 ATTGCTCATCAGTGGCCCGAGGGTCTCTGACAGTGGATGTCCTTTCTTGGAGAA 191  
 QY 750 AAACCTATCTGATCGTCCAGGACACCAAGATATGTCCTTGACAAAGGAGCTATACC 809  
 DB 192 AAACCTATCTGATCGTCCAGGACACCAAGATATGTCCTTGACAAAGGAGCTATACC 251  
 QY 810 CTAGTAAGCGGTTATCCGAAAGCGGCTGAGAAAGAAAGTGGAGACCCCTGATGGATATC 869  
 DB 252 CTAGTAAGCGGTTATCCGAAAGCGGCTGAGAAAGAAAGTGGAGACCCCTGATGGATATC 311  
 QY 870 CTGACTCTGTGAGATGCGGCTTTATCTGCTTCTGCTCCATATCATGAGCA 929  
 DB 312 CTGACTCTGTGAGATGCGGCTTTATCTGCTTCTGCTCCATATCATGAGCA 371  
 QY 930 GGACGGCGGCTGTGATGCTGAGACCTGAAGTCAAGAGCCCAAGCCAGTGAAGAGCTT 989  
 DB 372 GGACGGCGGCTGTGATGCTGAGACCTGAAGTCAAGAGCCCAAGCCAGTGAAGAGCTT 431  
 QY 990 CCTTGGCCCATGAGAAAGTACGAGAGCTTGTGTATGAAAGTCCCTTGGCCCTAAC 1049  
 DB 432 CCTTGGCCCATGAGAAAGTACGAGAGCTTGTGTATGAAAGTCCCTTGGCCCTAAC 491  
 QY 1050 TCATGTTCCGCCCAAGTGTCCCGCTTTTACTTCATTCATATGATGTCCTTACTCTAC 1109  
 DB 492 TCATGTTCCGCCCAAGTGTCCCGCTTTTACTTCATTCATATGATGTCCTTACTCTAC 551  
 QY 1110 AGTGATGTGAGAAACTGAATGACGCAAGG 1140  
 DB 552 AGTGATGTGAGAAACTGAATGACGCAAGG 582

RESULT 4  
 ABT08987  
 ID ABT08987 standard; DNA; 631 BP.  
 AC ABT08987;  
 XX 05-DEC-2002 (first entry)  
 DT Phase-1 Rat CT gene SEQ ID No 75.  
 XX Phase-1 Rat CT gene; da.  
 DE Rat; toxicity study; rat toxic response gene; toxicological response;  
 KW drug development; phase-1 rat CT gene; da.  
 XX Rattus sp.  
 OS MO200266682-A2.  
 PN 29-AUG-2002.  
 XX 29-JAN-2002; 2002WO-US02935.  
 PD 29-JAN-2001; 2001US-264933P.  
 XX 26-JUL-2001; 2001US-308161P.  
 PR (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.  
 PA Farris G, Hicken SH, Farr SB;  
 PI WPI; 2002-674961/72.  
 XX Evaluating the toxicity of an agent, useful in drug development or in  
 PT determining toxicological responses to a new drug, by determining the



PT New cell and tissue specific polynucleotides useful for diagnosis,  
PT prognosis or monitoring of treatments for disorders where the gene is  
PT associated with a cancer, immunopathology or neuropathology -

XX Claim 1; Page 244; 327pp; English.

CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide  
CC sequences (I). (I) can have cytostatic, immunomodulatory and  
CC neuroprotective activities, and can be used in gene therapy. (I) and  
CC proteins (II) encoded by them are used in high throughput screening  
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,  
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or  
CC their fragments, immunoglobulins, inhibitors, drug compounds and  
CC pharmaceutical agents. Expression of (I) in a sample indicates the  
CC differentiation of embryonic stem cells into a tissue selected from  
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic  
CC tissues. (I) and (II) are used to produce an expression profile that  
CC defines a metabolic or developmental process, treatment, condition,  
CC disease or disorder. The gene profile can be used for diagnosis,  
CC prognosis or monitoring of treatments and for investigating a  
CC predisposition to a disorder where the gene is associated with a  
CC cancer, immunopathology or neuropathology.

XX Sequence 1571 BP; 344 A; 431 C; 434 G; 362 T; 0 other;

Query Match 35.4%; Score 1126.8; DB 22; Length 1571;  
Best Local Similarity 86.3%; Pred. No. 0;  
Matches 1351; Conservative 0; Mismatches 2; Indels 213; Gaps 1;

QY 4 TCGAGCTCAGCATGCTAGGGTACTGAGAGACCCGTCATGCTGGGTTGTGAGCTAT 63  
DB 1 TCGAGCTCAGCATGCTAGGGTACTGAGAGACCCGTCATGCTGGGTTGTGAGCTAT 60  
QY 64 GCTGCTCTGCGCATTTGCCACCCCTCTTCTCCGACTAGTCCCATGGGAATGTTGCTG 123  
DB 61 GCTGCTCTGCGCATTTGCCACCCCTCTTCTCCGACTAGTCCCATGGGAATGTTGCTG 120  
QY 124 AAGGCGAGACCAAGCCAGACCCGAGCTGCTGAAGCTGCTGAGTGGCTGAGGCTTG 183  
DB 121 AAGGCGAGACCAAGCCAGACCCGAGCTGCTGAAGCTGCTGAGTGGCTGAGGCTTG 180  
QY 184 ATGCTACCACTCGATGATGACATGCAATGCTGTTTTTAAAGGGAATTTGTGTGA 243  
DB 181 ATGCTACCACTCGATGATGACATGCAATGCTGTTTTTAAAGGGAATTTGTGTGA 240  
QY 244 AGAGTCACAATGGGACCGGAGTTATCTCAAGAGATGGAATTTCCCAAGCCCTG 303  
DB 241 AGAGTCACAATGGGACCGGAGTTATCTCAAGAGATGGAATTTCCCAAGCCCTG 300  
QY 304 TGGATGCTGATCCGTCAGAGTCAACAAGTCTTTTTCGATCAAGGGGGAACAAGTCT 363  
DB 301 TGGATGCTGATCCGTCAGAGTCAACAAGTCTTTTTCGATCAAGGGGGAACAAGTCT 360  
QY 364 GGGTATACCTCTCTGAAAAGAGAGAAAGATACCAAGTTGCTCAAGATGAATTTT 423  
DB 361 GGGTATACCTCTCTGAAAAGAGAGAAAGATACCAAGTTGCTCAAGATGAATTTT 420  
QY 424 CTGGAAATCCCATCCCACTGAGATGCAAGCTGTGAATGTCAACGTGAAGATGCAAGCTG 483  
DB 421 CTGGAAATCCCATCCCACTGAGATGCAAGCTGTGAATGTCAACGTGAAGATGCAAGCTG 480  
QY 484 AAGGCGCTCGCTTTCTTCCA----- 502  
DB 481 AAGGCGCTCGCTTTCTTCCAAGGATGACCGGAGTGGTTCTGGGACTTGGCTACCGGAAACA 540  
QY 503 ----- 502  
DB 541 TGAAGAGAGGTTCTGGCCAGCTGTTGGGAACGTCTCTGCTGAGATGGCTGGGCC 600  
QY 503 ----- 502  
DB 601 GCTACTACTGCTTCAAGGATCAACAATCTCTGCGCTTCAAGCTGTCAGGAGAGAGTGC 660

QY 503 -----AGGCCATG 510  
DB 661 CTCACAGTACCCCGGAGATGTCGAGACTACTTCAATGCTCCCTCGGACAGAGGCTATG 720  
QY 511 GACACAGAAATGGGACTGGCCATGGGAACGATACCAACATGAGCCCTGATATATGCGCT 570  
DB 721 GACACAGAAATGGGACTGGCCATGGGAACGATACCAACATGAGCCCTGATATATGCGCT 780  
QY 571 GTAGCCCAATATGATCTGTGTGCACTGACGTGACAAACCATGCTGACACTATGCT 630  
DB 781 GTAGCCCAATATGATCTGTGTGCACTGACGTGACAAACCATGCTGACACTATGCT 840  
QY 631 TAGTGGGACCACTACTGCGCTTGGACACCAAGCCGAGATGAGTATGCTGAGCCCA 690  
DB 841 TAGTGGGACCACTACTGCGCTTGGACACCAAGCCGAGATGAGTATGCTGAGCCCA 900  
QY 691 TTGCTCATGATGAGCCCAAGGCTCTTCAAGATGATGCTGCTTCTTCTGGAAGAAA 750  
DB 901 TTGCTCATGATGAGCCCAAGGCTCTTCAAGATGATGCTGCTTCTTCTGGAAGAAA 960  
QY 751 AACTCTATGCTGCTGAGGACACCCAGATATATGCTTCTGACAAAGGAGCTATACCC 810  
DB 961 AACTCTATGCTGCTGAGGACACCCAGATATATGCTTCTGACAAAGGAGCTATACCC 1020  
QY 811 TAGTAAGCGGTTATCCGAGCGGCTGAGAGAGAAATGCGGACCCCTCATGGGATATCC 870  
DB 1021 TAGTAAGCGGTTATCCGAGCGGCTGAGAGAGAAATGCGGACCCCTCATGGGATATCC 1080  
QY 871 TGAATCTGTGATGAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 930  
DB 1081 TGAATCTGTGATGAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
QY 931 GACGCGGCTGT 990  
DB 1141 GACGCGGCTGT 1200  
QY 991 CTTGGCCCATGAGAAAGTGAAGCGAGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1050  
DB 1201 CTTGGCCCATGAGAAAGTGAAGCGAGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260  
QY 1051 CATGTTCCGCAATGATGCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1110  
DB 1261 CATGTTCCGCAATGATGCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
QY 1111 GTGATGTGAGAAATGGAATGACAGCCAGCCCTTCCGGAACCCGGAATGTGACAGATC 1170  
DB 1321 GTGATGTGAGAAATGGAATGACAGCCAGCCCTTCCGGAACCCGGAATGTGACAGATC 1380  
QY 1171 TCTGGGCTGCACTCACTGAGGGGCTTCTGACATGATCTGGCTGGGCCCACTCTTA 1230  
DB 1381 TCTGGGCTGCACTCACTGAGGGGCTTCTGACATGATCTGGCTGGGCCCACTCTTA 1440  
QY 1231 GTTCTCATATTAAGACAGATGCTTCTTGTCTTCACTGAGGGGCTTCTGACATGA 1290  
DB 1441 GTTCTCATATTAAGACAGATGCTTCTTGTCTTCACTGAGGGGCTTCTGACATGA 1500  
QY 1291 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350  
DB 1501 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
QY 1351 AATCAA 1356  
DB 1561 AAAAAA 1566

RESULT 2  
ID AAH98592  
AAH98592 standard; cDNA; 1610 BP.

AC AAH98592;  
AC AAH98592;  
DT 12-OCT-2001 (first entry)

